DRAFT BACKGROUND REVIEW DOCUMENT

IN VITRO ACUTE TOXICITY TEST METHODS

National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)

March 17, 2006

TABLE OF CONTENTS

APPENDICES

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

A-CUTE-TOX	A-Cute-Tox Project (EU Research & Development Integrated Project)
ASTDR	Agency for Toxic Substances and Disease Registry
ASTM	American Society for Testing and Materials
ATC	Acute Toxicity Class
ATCC	American Type Culture Collection
BBB	Blood:brain barrier
BRD	Background Review Document
CAS	Chemical Abstracts Service
CASRN	Chemical Abstracts Service Registry Number
CPSC	U.S. Consumer Product Safety Commission
CTFA	Cosmetic, Toiletries and Fragrance Association
CV	Coefficient of Variation
$^{\circ}$C	Degrees Celsius
DOD	U.S. Department of Defense
DOT	U.S. Department of Transportation
EC	European Commission
EC/HO	European Commission/British Home Office
ECETOC	European Centre for Ecotoxicology and Toxicology Of Chemicals
ECVAM	European Center for the Validation of Alternative Methods
EDIT	Evaluation-guided development of new *in vitro* tests
EPA	U.S. Environmental Protection Agency
EU	European Union
FAL	FRAME Alternatives Laboratory
FDA	U.S. Food and Drug Administration
FR	Federal Register

FRAME	Fund for the Replacement of Animals in Medical Experiments
GHS	Globally Harmonized System
GLP	Good Laboratory Practice
HSDB	Hazardous Substances Database
IC$_{50}$	Inhibitory concentration producing 50% inhibition of the endpoint measured
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
IIVS	Institute for In Vitro Sciences
INVITOXX	*In Vitro* Techniques in Toxicology (ERGATT FRAME ECVAM Data bank)
IRAG	Interagency Regulatory Alternatives Group
K$_{ow}$	Octanol-Water Partition Coefficient
LC	Lethal concentration
LD$_{50}$	Dose that produces lethality in 50% of test animals
MEIC	Multicentre Evaluation of *In Vitro* Cytotoxicity
NCS	Newborn calf serum
NHK	Normal human epidermal keratinocytes
NICEATM	National Toxicology Program Center for the Evaluation of Alternative Toxicological Methods
NIEHS	National Institute of Environmental Health Sciences
NIH	National Institutes of Health
NIOSH	U.S. National Institute for Occupational Safety and Health
NLM	National Library of Medicine
NR	Neutral red
NRU	Neutral red uptake
NTP	U.S. National Toxicology Program
OD	Optical density

OECD	Organisation for Economic Cooperation and Development
OPPTS	EPA Office of Prevention, Pesticides and Toxic Substances
OSHA	U.S. Occupational Safety & Hazards Administration
PC	Positive control
QA	Quality Assurance
RC	Registry of Cytotoxicity
RTECS	Registry of Toxic Effects for Chemical Substances
SD	Standard deviation
SMT	Study Management Team (NICEATM/ECVAM validation study)
3T3	BALB/c mouse fibroblasts, clone A31 (ATCC # CCL-163)
TSCA	Toxic Substances Control Act
UDP	Up-and-Down Procedure
UN	United Nations
VC	Vehicle control
WHO	World Health Organization
ZEBET	German Center for Documentation and Evaluation of Alternative Methods to Animal Experiments

ACKNOWLEDGEMENTS

The following individuals are acknowledged for their contributions to the in vitro acute toxicity test method review process.

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) Acute Toxicity Working Group (ATWG)

Consumer Product Safety Commission (CPSC)

Patricia Brundage, Ph.D.
Kailash Gupta, Ph.D.
Cassandra Prioleau, Ph.D.
Marilyn Wind, Ph.D. (ICCVAM Co-Chair)

Department of Energy (DOE)

Po-Yung Lu, Ph.D.

Environmental Protection Agency (EPA)

Karen Hamernik, Ph.D.
Masih Hashim, Ph.D.
Marianne Lewis, Ph.D.
Elizabeth Margosches, Ph.D.
Debbie McCall, Ph.D.
John Redden, Ph.D.
Amy Rispin, Ph.D.

Food and Drug Administration (FDA)

Leonard Schechtman, Ph.D. (ICCVAM Chair)
Kenneth Hastings, Ph.D.
Abigale Jacobs, Ph.D.
Suzanne Morris, Ph.D.
David Morse, Ph.D.
Thomas Umbreit, Ph.D.

National Institute for Occupational Safety & Health (NIOSH)

Stephen Reynolds, Ph.D.

National Institute of Environmental Health Sciences (NIEHS)

Rajendra Chhabra, Ph.D.
William Stokes, D.V.M., D.A.C.L.A.M. (Executive Director, ICCVAM)
Raymond Tice, Ph.D., (Deputy Director, NICEATM)

National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)

Participants in the *In Vitro* Cytotoxicity Validation Study

NICEATM gratefully acknowledges the generous contributions of the individuals who directly participated in the NICEATM/ECVAM Validation Study. Their time and efforts are greatly appreciated.

BioReliance Corp.
(Chemical Distribution)
Dr. Martin Wenk – Principal Investigator

Institute for In Vitro Sciences (IIVS)
(Lead Laboratory – Protocols)
Mr. Hans Raabe – Study Director
Mr. Greg Mun – Laboratory Manager
Ms Angela Sizemore – Research Technician
Mr. Gregory O. Moyer – Research Technician
Dr. John Harbell – Scientific Director

U.S. Army Edgewood Chemical and
Biological Center (ECBC)
(Testing Laboratory)
Dr. Cheng Cao – Study Director
Ms Janna Madren-Whalley – Research Technician
Dr. Chundakkadu Krishna – Research Technician
Dr. James J. Valdes – Scientific Advisor

FRAME Alternatives Laboratory (FAL)
University of Nottingham, UK
(Lead Laboratory – Software)
Dr. Richard Clothier – Study Director
Ms Nicola Bourne – Research Technician
Ms Monika Owen – Research Technician
Ms Rachel Budworth – Research Technician

Constella Group
(Statistical Analyses)
Dr. Patrick Crockett
Dr. Eric Harvey
Mr. Robert Lee
Ms Janine Wilcox

Private Statistical Consultant
Dr. Joseph Haseman

NIEHS
Dr. Grace Kissling – Contract Project Officer
Ms Molly Vallant - Contract Project Officer

Study Management Team (SMT)
NICEATM
Dr. William Stokes (NIEHS) – Director
Dr. Raymond Tice (NIEHS) – Asst. Director – Advisor
Dr. Judy Strickland (ILS, Inc.) – Project Coordinator
Mr. Michael Paris (ILS, Inc.) – Asst. Project Coordinator
Dr. Jeffrey Charles (ILS, Inc.) – Advisor

ECVAM
Dr. Thomas Hartung – Head of Unit
Dr. Silvia Casati – Task Leader

PREFACE

The Institute of Medicine estimates that more than 4 million poisonings occur annually in the United States (Institute of Medicine 2004). In 2001, 30,800 deaths placed poisoning as the second leading cause of injury-related death behind automobile accidents (42,433 deaths) (Institute of Medicine 2004). In order to ensure that all potentially hazardous substances have proper warning labels, regulatory agencies require determination of acute toxicity hazard potential of substances and products. This determination for oral acute toxicity hazard is currently made using a test that requires laboratory rats. Historically, lethality estimated by the LD_{50} (i.e., the dose of a test substance that produces death in 50% of the animals tested) has been a primary toxicological endpoint in acute toxicity tests.

The conventional LD_{50} acute oral toxicity *in vivo* test method has been modified in various ways to reduce and refine[1] animal use in toxicity testing (OECD 2001a, c, d; EPA 2002a). Most recently, the LD_{50} was replaced, for hazard classification testing purposes, with the UDP, based on an Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) technical evaluation and formal ICCVAM recommendations (ICCVAM 2000, 2001c). This method now reduces animal use by over 70% compared to the previous method.

In 1999, at the request of the U.S. Environmental Protection Agency (EPA) Office of Pesticides, Prevention, and Toxic Substances, ICCVAM reviewed the validation status of *in vitro* methods for estimating acute oral toxicity. This request was based on studies published in recent years that showed a correlation between *in vitro* and *in vivo* acute toxicity. *In vitro* cytotoxicity methods have been evaluated as another means to reduce and refine the use of animals and these methods may be helpful in predicting *in vivo* acute toxicity. Since moving the starting dose closer to the LD_{50} reduces the number of animals necessary for the acute

---

[1] A reduction alternative is a new or modified test method that reduces the number of animals required. A refinement alternative is a new or modified test method that refines procedures to lessen or eliminate pain or distress in animals or enhances animal well-being (ICCVAM 2003).

oral systemic toxicity test, the use of *in vitro* cytotoxicity assays to predict a starting dose close to the LD_{50} may reduce animal use.

In October of 2000, the International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity sponsored by the National Toxicology Program (NTP), the National Institute of Environmental Health Sciences (NIEHS) and the EPA was convened in Arlington, VA. The Organizing Committee invited 33 expert scientists from academia, industry, and government agencies to participate in the Workshop. Invited scientific experts and ICCVAM agency scientists were assigned to one of four Breakout Groups and prepared recommendations on the following:

- *In Vitro* Screening Methods for Assessing Acute Toxicity
- *In Vitro* Methods for Toxicokinetic Determinations
- *In Vitro* Methods for Predicting Organ Specific Toxicity
- Chemical Data Sets for Validation of *In Vitro* Acute Toxicity Test Methods

Workshop participants concluded that none of the proposed *in vitro* methods had been formally evaluated for reliability and relevance, and that their usefulness and limitations for generating information to meet regulatory requirements for acute toxicity testing had not been adequately assessed. However, an *in vitro* approach proposed by the German Center for Documentation and Evaluation of Alternative Methods to Animal Experiments (ZEBET) was recommended for rapid adoption so that data could be generated to establish its usefulness with a large number of chemicals (ICCVAM 2001a). In addition, a separate *Guidance Document on Using In Vitro Data to Estimate In Vivo Starting Doses for Acute Toxicity* (ICCVAM 2001b) was prepared to provide sample cytotoxicity protocols and instructions for using *in vitro* data to predict starting doses for acute *in vivo* systemic toxicity tests.

ICCVAM, which is charged with coordinating the technical evaluations of new, revised, and alternative test methods with regulatory applicability (ICCVAM Authorization Act of 2000, Public Law 106-545; available: http://iccvam.niehs.nih.gov/about/PL106545.pdf), agreed that *in vitro* basal cytotoxicity test methods should have a high priority for evaluation. The

National Toxicology Program (NTP) Center for the Evaluation of Alternative Toxicological Methods (NICEATM) collaborated with the European Center for the Validation of Alternative Methods (ECVAM), a component of the European Commission's Joint Research Centre, to further characterize the usefulness of *in vitro* cytotoxicity assays as predictors of starting doses for acute oral lethality assays. NICEATM and ECVAM designed a multi-laboratory validation study to evaluate the performance of two standardized *in vitro* basal cytotoxicity test methods using 72 reference substances with the ZEBET approach of using the Registry of Cytotoxicity (RC) regression model. Based on the procedures described in the *Guidance Document* (ICCVAM 2001b), the validation study used two mammalian cell types (i.e., BALB/c 3T3 mouse fibroblasts [3T3] and a primary normal human epidermal keratinocytes [NHK]) for *in vitro* basal cytotoxicity test methods with a neutral red uptake (NRU) cell viability endpoint to predict starting for acute oral systemic toxicity test methods. The inclusion of human cells in the validation study also implements another workshop recommendation, that of evaluating whether cytotoxicity in human or rodent cells can be used to predict human acute toxicity.

The objectives identified for the validation study were to:
- further standardize and optimize two *in vitro* NRU cytotoxicity protocols using 3T3 cells or NHK cells in order to maximize intra- and inter-laboratory reproducibility
- refine the prediction model drawn from the ZEBET approach
- assess the accuracy of the two standardized *in vitro* basal cytotoxicity test methods for estimating rodent oral LD_{50} values across the five Globally Harmonized System of Classification and Labelling of Chemicals (GHS; UN 2005) categories of acute oral toxicity as well as unclassified toxicities and estimating human lethal serum concentrations
- estimate the reduction and refinement in animal use achievable from using *in vitro* basal cytotoxicity assays as one of the factors of the weight-of-evidence to identify starting doses for specific *in vivo* acute toxicity tests

- generate high quality *in vivo* lethality and *in vitro* cytotoxicity databases that can be used to support the investigation of other *in vitro* test methods necessary to improve the prediction of acute systemic toxicity

Scientists assembled for the ICCVAM-sponsored scientific peer review panel meeting ("Panel") on May 23, 2006 will independently assess the usefulness and limitations of the *in vitro* basal cytotoxicity test methods to predict starting doses for acute oral systemic toxicity test methods. The Background Review Document (BRD) on the two *in vitro* NRU test methods prepared by NICEATM and provided to the peer review panel and the public contains:

1. comprehensive summaries of the data generated in the validation study
2. an analysis of the accuracy and reliability of the test method protocols
3. related information characterizing the potential animal savings produced by using the *in vitro* basal cytotoxicity test methods as adjuncts to specific acute systemic toxicity test methods

The Panel will also evaluate draft test method performance standards, protocols, and draft ICCVAM recommendations. The public is invited to provide comments on the BRD and other documents and to attend the Panel meeting. Prior to this meeting, any public comments provided about the documents will be provided to the Panel for their consideration. The BRD can be obtained from the ICCVAM/NICEATM Web site (http://iccvam.niehs.nih.gov) or by contacting NICEATM.

Following the conclusion of the Panel meeting, the ICCVAM and its Acute Toxicity Working Group (ATWG) will consider the Panel report, the performance standards for the use of *in vitro* basal cytotoxicity test methods to predict starting doses for acute systemic toxicity test methods, and any public comments in preparing its final test method recommendations for these *in vitro* basal cytotoxicity test methods. These recommendations will be made available to the public and provided to the U.S. Federal agencies for consideration, in accordance with the ICCVAM Authorization Act of 2000 (Public Law 106-545).

On behalf of the ICCVAM, we gratefully acknowledge the many contributions of all who participated in the *in vitro* cytotoxicity validation study and those who assisted in the preparation of the documents evaluated at the peer review meeting. We extend a special thanks to the participating laboratory Study Directors and scientists who worked diligently to provided critical data and information. We also thank the ECVAM scientists who participated in the management of the validation study and who provided valuable information, comments, and opinions throughout the study. The efforts of the ATWG members were instrumental in assuring a complete and informative BRD. The efforts of the NICEATM staff in coordinating the validation study, providing timely distribution of information, and preparing the various documents are acknowledged and appreciated. We especially acknowledge Dr. Judy Strickland and Mr. Michael Paris for their coordination of the validation study and preparation of the BRD and other documents.

William S. Stokes, D.V.M. Diplomate A.C.L.A.M.
Director, NICEATM
Executive Director, ICCVAM

Leonard Schechtman, Ph.D.
U.S. Food and Drug Administration
National Center for Toxicological Research
Chairman, ICCVAM

March 17, 2006

EXECUTIVE SUMMARY

This Background Review Document (BRD) describes the results of a validation study conducted to characterize two *in vitro* basal cytotoxicity tests for determining starting doses for acute oral systemic toxicity assays. The purpose of these tests is to reduce the total number of animals needed for in the *in vivo* tests. As part of this study, methods for two *in vitro* neutral red uptake (NRU) assays using mouse fibroblast (BALB/c) 3T3 cells or normal human epidermal keratinocytes (NHK) were standardized and optimized, the accuracy and validity of the tests were determined using reference chemicals of various toxicities, and computer simulation models were used to estimate the potential reduction in animal usage that could be accomplished by the use of these assays. In addition, high quality *in vivo* lethality and *in vitro* cytotoxicity databases were generated that may be useful in other validation studies for *in vitro* toxicity tests.

The results of the study showed that the 3T3 and NHK NRU test methods are not sufficiently accurate as stand-alone methods to correctly predict acute oral toxicity. However, based on computer simulations for the reference substances tested in this study, the use of these *in vitro* basal cytotoxicity test methods for the selection of starting doses for *in vivo* testing has the potential to reduce both the numbers of animals needed and animal deaths compared to the default procedures.

Introduction and Rationale

Although *in vitro* basal cytotoxicity test methods are not currently regarded as suitable replacements for acute oral systemic toxicity assays (Spielmann et al. 1999; ICCVAM 2001a), such test methods have been evaluated as a means to reduce and refine[2] the use of animals in acute oral systemic toxicity testing. In 1983, an international effort, the Multicentre Evaluation of *In Vitro* Cytotoxicity (MEIC), was initiated to evaluate the relationship of *in vitro* cytotoxicity to acute *in vivo* toxicity. Tests of 50 substances in 61 *in*

---

[2] A reduction alternative is a new or modified test method that reduces the number of animals required. A refinement alternative is a new or modified test method that refines procedures to lessen or eliminate pain or distress in animals or enhances animal well-being (ICCVAM 2003).

vitro assays identified a battery of three human cell line assays that were highly correlated to human lethal blood concentrations. The Registry of Cytotoxicity (RC), a database of 347 substances that currently consists of *in vivo* acute toxicity data from rats and mice and *in vitro* cytotoxicity data from multiple cell lines, was published in 1998 (Halle 1998). A regression formula (the RC millimole regression) constructed from these data was proposed by ZEBET, the German National Center for the Documentation and Evaluation of Alternative Methods to Animal Experiments, as a method to reduce animal use by identifying the most appropriate starting doses for acute oral systemic toxicity tests (Halle 1998; Spielmann et al. 1999). These initiatives (and others) to use *in vitro* cytotoxicity test methods to reduce animal use in acute toxicity testing were evaluated by the International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity in October 2000 ("Workshop 2000"; ICCVAM 2001a). This workshop was organized by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM).

ICCVAM recommended (ICCVAM 2001a) further evaluation of the use of *in vitro* cytotoxicity data as one of the factors used to estimate starting doses for *in vivo* acute lethality studies based on preliminary information that this approach could reduce the number of animals used in *in vivo* studies (i.e., reduction), minimize the number of animals that receive lethal doses (i.e., refinement), and avoid underestimating hazard. To assist in the adoption and implementation of the ZEBET approach, the *Guidance Document on Using In Vitro Data to Estimate In Vivo Starting Doses for Acute Toxicity* (hereafter referred to as *Guidance Document*; ICCVAM 2001b) was prepared by ICCVAM with the assistance of several workshop participants.

ICCVAM concurred with the Workshop 2000 recommendation that near-term validation studies should focus on two standard basal cytotoxicity assays: one using a human cell system and one using a rodent cell system. Historical data for *in vitro* cytotoxicity testing using 3T3 cells is available through other publications (e.g., Balls et al. 1995; Brantom et al. 1997; Gettings et al. 1991, 1994a, 1994b; Spielmann et al. 1991, 1993, 1996). Historical data

for *in vitro* basal cytotoxicity testing using normal human keratinocytes (NHK) cells are also available through other publications (e.g., Gettings et al. 1996; Harbell et al. 1997; Sina et al. 1995; Willshaw et al. 1994).

NICEATM, in partnership with the European Center for the Validation of Alternative Methods (ECVAM), designed a multi-laboratory validation study to evaluate reduction or refinement that might result when using cytotoxicity data from the 3T3 and NHK NRU test methods as part of the weight-of-evidence to estimate starting doses for the Up-and-Down Procedure (UDP; OECD 2001a; EPA 2002a) and the Acute Toxic Class (ATC) method (OECD 2001d). The *Guidance Document* NRU protocols were the initial basis of the NICEATM/ECVAM study protocols. These protocols were derived from examination of the BALB/c 3T3 Cytotoxicity Test, INVITTOX Protocol No. 46 (available at the FRAME-sponsored INVITTOX database [http://embryo.ib.amwaw.edu.pl/invittox/]) and the Borenfreund and Puerner (1985) (3T3 cells) as well as Borenfreund and Puerner (1984) and (Heimann and Rice 1983) (NHK cells). See **Section 2** for a detailed description of the test method protocols.

Test Method Protocol Components
The test method protocol components for the *in vitro* NRU cytotoxicity test methods used in the NICEATM/ECVAM study are very similar for the 3T3 and the NHK cells. The following procedures are common to both cell types:

- preparation of reference substances and positive control
- cell culture environmental conditions
- determination of test substance solubility
- 96-well plate configuration for testing samples
- range finder and definitive testing (48-hour exposure to the reference substance)
- microscopic evaluation of cell cultures for toxicity
- measurement of NRU
- data analysis

The main differences in the test methods are:

- the conditions of propagation of the cells in culture

- the cell growth medium components

- the application of reference substances to the 96-well plate (i.e., different volumes of reference substance solution)

Three testing laboratories participated in testing 72 reference substances, in three phases:

- ECBC: The U.S. Army Edgewood Chemical Biological Center (Edgewood, MD)

- FAL: Fund for the Replacement of Animals in Medical Experiments (FRAME) Alternatives Laboratory (Nottingham, UK)

- IIVS (Gaithersburg, MD)

BioReliance Corporation (Rockville, MD) procured and distributed the coded reference substances and performed solubility tests prior to distribution to the cytotoxicity testing laboratories.

Validation Substances

Reference substances were selected to represent: (1) the complete range of *in vivo* acute oral toxicity (in terms of LD_{50} values where LD_{50} is median lethal dose); (2) the types of substances regulated by various regulatory authorities; and (3) those with human toxicity data and/or human exposure potential. To assure the complete range of toxicity was covered, the Globally Harmonized System of Classification and Labelling of Substances (UN 2005) was used to select 12 substances for each of the five acute oral toxicity categories and 12 unclassified substances. A discussion of characteristics and sources of the reference substances can be found in **Section 3** of the BRD. The set of selected reference substances had the following characteristics:

- 58 of the 72 substances were also included in the RC

- 27 (38%) of the substances had pharmaceutical uses, 15 (21%) had pesticide uses, 8 (11%) had solvent uses, and 5 (7%) had food additive uses. The

   remaining substances were used for a variety of manufacturing and consumer
   products.

- 55 (76%) were organic compounds and 17 (24%) were inorganic compounds;
  commonly represented classes of organic compounds included heterocyclic
  compounds, carboxylic acids, and alcohols

- 22 (31%) substances were known or expected to have active metabolites

- many of the selected substances had multiple target organs/effects; including
  neurological, liver, kidney, and cardiovascular effects

In Vivo Rodent Toxicity Reference Data

Because the *in vitro* NRU cytotoxicity test methods are intended to be used as adjuncts to *in
vivo* acute oral systemic toxicity test methods using rats, rodent LD_{50} values from acute oral
systemic toxicity tests are the most appropriate reference data for the *in vitro* NRU IC_{50}
values (i.e., the concentration of the test substance that reduces cell viability by 50%). *In
vivo* LD_{50} reference data for the 72 reference substances were determined from the literature.
Limiting the data to studies conducted under Good Laboratory Practice (GLP) guidelines
(OECD 1998; EPA 2003a, 2003b; FDA 2003) was not possible since only 3% of the data
records were from such studies. While mouse data were considered initially, eventually
analyses were restricted to rat data. In total, 485 acute oral LD_{50} values were identified for
rats for the 72 reference substances. Reference LD_{50} values for each substance were
identified by excluding studies that employed the following materials and methods:

- feral rats

- rats < 4 weeks of age

- anesthetized rats

- test substance administered in food or capsule

- LD_{50} reported as a range or inequality

In vivo reference values were determined, where multiple values existed, by calculating a
geometric mean of the values. The reference LD_{50} values for 20 of the 72 substances varied

enough from the initial LD_{50} values, which came from the RC database and other summary sources, that the substances were reclassified into different GHS oral toxicity categories.

Test Method Accuracy

Although the 3T3 and NHK NRU test methods are not intended as replacements for acute systemic toxicity assays, the ability of these methods to correctly predict the reference LD_{50} values was used to evaluate their accuracy[3]. The rationale for evaluating the accuracy of LD_{50} predictions was that the animal savings produced by using these *in vitro* test methods to predict starting doses for acute systemic toxicity assays would be greatest when the starting dose is as close as possible to the LD_{50}. An IC_{50}-LD_{50} regression model was used to derive the estimated LD_{50} value using 3T3 or NHK NRU IC_{50} values.

A number of different analyses were done in an attempt to improve the estimation of LD_{50} by the regression. IC_{50}-LD_{50} regressions (millimole units) for each NRU test method and laboratory were developed using the IC_{50} data and reference LD_{50} for the reference substances in the NICEATM/ECVAM validation study. The regressions were not significantly different from a regression for the 58 RC substances (calculated using the RC IC_{50} and LD_{50} data) included among the 72 reference substances (F test; p = 0.929 for the 3T3 NRU regression and p = 0.144 for the NHK NRU regression).

Discordant substances (i.e., test substances that fit the RC millimole regression poorly) were evaluated. Since the 3T3 and NHK NRU regressions yielded results that were not different from the RC, the RC millimole regression was preferred for analysis of discordant substances since it is based on a larger chemical data set than that used in the NICEATM/ECVAM validation study. Discordant substances from the NICEATM/ECVAM study were analyzed to determine whether there were relationships between their outlier status and physical or chemical characteristics. The lack of fit to the RC millimole regression was correlated with chemical class, boiling point, molecular weight, and log K_{OW}, but not with the insolubility of

---

[3] Accuracy: the agreement between a test method result and an accepted reference value (ICCVAM 2003).

the reference substance in the 3T3 or NHK medium or to the fact that the test method systems had little to no metabolic capability. Since these test methods are based upon basal cytotoxicity, mechanism of toxicity was also considered as a characteristic to explain poor fit to the RC millimole regression. Of the 21 reference substances with specific mechanisms of toxicity that were not expected to be active in the 3T3 and NHK cell cultures, 13 (62%) were outliers (i.e., they fit the RC millimole regression poorly). These substances represented 13/30 (43%) of the outliers for the 3T3 NRU and 13/31 (42%) for the NHK NRU. Information on this analysis is presented in **Section 6.4**.

Additional regressions were developed to improve the RC millimole regression. Substances with *in vivo* LD_{50} values based only on mouse test data were excluded. Substances with mechanisms of toxicity that were not expected to be active in the 3T3 and NHK cell cultures were excluded, leading to the RC rat-only regression excluding substances with specific mechanisms of toxicity. In addition, the RC rat-only data were converted to a weight basis for an additional regression analysis, the RC rat-only weight regression.

Accuracy of the *in vitro* NRU test methods (when used with each of the three IC_{50}-LD_{50} regressions) was characterized by determining the proportion of chemicals for which GHS acute oral toxicity categories were correctly predicted. However, this does not imply that the *in vitro* NRU tests are stand-alone methods that can be used for hazard classification. The accuracy for the prediction of toxicity for substances in the GHS acute oral toxicity categories for $LD_{50} > 2000$ mg/kg was improved by removing substances with specific mechanisms of toxicity from the RC rat-only weight regression (compared with the RC millimole regression). It did not improve the accuracy of category prediction for substances with $LD_{50} < 50$ mg/kg or for substances with $300 < LD_{50} \leq 2000$ mg/kg; however, in the latter case, accuracy was already relatively high. The RC rat-only weight regression excluding substances with specific mechanisms of toxicity improved the overall accuracy for the 3T3 NRU test method from 26% (12/46 test substances) with the RC millimole regression to 46% (21/46 test substances). The RC rat-only weight regression excluding substances with specific mechanisms of toxicity improved the overall accuracy for the NHK NRU test method from 28% (13/47 test substances) for the RC millimole regression to 38%

(18/47 test substances). For each regression evaluated, there was a general trend to underpredict the toxicity of the most toxic chemicals and to overpredict the toxicity of the least toxic chemicals. A detailed discussion of the accuracy analyses is presented in **Section 6.3.**

Test Method Reliability

Intra- and inter-laboratory reproducibility of the 3T3 and NHK NRU IC_{50} data were assessed using analysis of variance (ANOVA), coefficient of variation (CV) analysis, comparison of the laboratory-specific IC_{50}-LD_{50} regressions to one another (for each test method), and laboratory concordance for the GHS acute oral toxicity category predictions. Reproducibility is the consistency of individual test results obtained in a single laboratory (intralaboratory reproducibility) or in different laboratories (interlaboratory reproducibility) using the same protocol and test samples.

Although ANOVA results for the positive control, SLS, IC_{50} for the 3T3 NRU test method indicated there were significant differences among laboratories (p = 0.006), a graphical display of the data (see **Figure 7-1**) shows that laboratory means and standard deviations for each study phase overlap one another. Interlaboratory CV values, which ranged from 2% to 10% for the study phases, also indicated that the laboratories were similar. ANOVA results for the SLS IC_{50} for the NHK NRU test method also showed significant differences between laboratories (p < 0.001). A different cell culture method at FAL was responsible for SLS IC_{50} differences among the laboratories in Phases Ia and Ib. After harmonization of culture methods with the other laboratories, the laboratory means and standard deviations were quite similar for Phases II and III (see **Figure 7-1**). Interlaboratory CV values for SLS in the NHK NRU test method ranged from 8% (Phase III) to 39% (Phase Ia). Very small slopes (< |0.001|) for linear regression analyses of the SLS IC_{50} over time (within each laboratory) for both *in vitro* NRU test methods indicated that the SLS IC_{50} was stable over the 2.5 year duration of the study.

ANOVA results for the reference substances showed significant laboratory differences

for 26 substances for the 3T3 NRU test method and seven substances for the NHK NRU test method (see **Table 7-6**). An analysis to determine the relationship, if any, between substance attributes and interlaboratory CV indicated that physical form, solubility, and volatility had little effect on CV. CV seemed to be related, however, to chemical class, GHS acute toxicity category, IC_{50}, and boiling point (see **Section 7.2.2**). Although the ANOVA results and the interlaboratory CV analysis (at least for the 3T3 NRU) seemed to indicate that interlaboratory reproducibility may be less than desired, the comparison of laboratory specific IC_{50}-LD_{50} regressions indicated that the laboratory regressions for both test methods were not significantly different from one another (p = 0.796 for the 3T3 NRU and p = 0.985 for the NHK NRU). In addition, the laboratory concordance for the prediction of GHS oral toxicity categories ranged from 78 - 85% for the 3T3 NRU and 84 - 91% for the NHK NRU (depending on the regression used). The similarity of the laboratories in LD_{50} predictions (via regression) and GHS toxicity category predictions is considered most significant with respect to the reproducibility analyses since the NRU methods are proposed for use with the regressions in determining starting doses for acute oral toxicity tests.

Animal Welfare Considerations: Reduction, Refinement, and Replacement
For the NICEATM/ECVAM validation study, computer simulation models were used to simulate the UDP and ATC testing of the reference substances tested with the NRU basal cytotoxicity test methods. Reference substances that had only mouse reference LD_{50} data or with known mechanisms of toxicity that were not expected to be active in the 3T3 and NHK cell cultures were not evaluated. The number of animals used for simulated testing and the number of animals that lived or died were determined for the default starting dose and for the NRU-determined starting dose (i.e., one default dose lower than the estimated LD_{50}) with 2000 computer test simulations for each substance and starting dose. The computer simulations accounted for the accuracy of the NRU results with respect to the prediction of LD_{50} values since the accuracy was conferred by the particular regression evaluated

Computer simulation modeling of UDP testing shows that, for the substances tested in this validation study, the prediction of starting doses using the NRU test methods resulted in the use of statistically fewer animals by an average of 1.00 - 1.16 animals (approximately 12%)

when using the RC rat-only weight regression excluding substances with specific mechanisms of toxicity depending upon NRU test method and dose-response slope (of 2 or 8.3). There were no animal savings for chemicals with $50 < LD_{50} \leq 300$ mg/kg when test substances were grouped by GHS toxicity category since animal use was compared with animal used for the default starting dose of 175 mg/kg. However, statistically significant animal savings were as high as 1.75 - 2.22 (19.1 - 20.5%) animals for substances with $2000 < LD_{50} \leq 5000$ mg/kg or $LD_{50} > 5000$ mg/kg. Using the NRU test methods to estimate starting doses also resulted in approximately 0.1 to 0.2 fewer deaths for the simulated UDP testing compared to the default starting dose.

Computer simulation modeling of ATC testing showed that, for the substances tested in this validation study, the prediction of starting doses using the NRU test methods resulted in the use of 1.68 - 1.94 (15.4 - 21.1%) fewer (statistically) animals for the RC rat-only weight regression excluding substances with specific mechanisms of toxicity depending upon NRU test method and dose-response slope (of 2 or 8.3). There were no animal savings for substances with $300 < LD_{50} \leq 2000$ mg/kg when test substances were grouped by GHS toxicity category since animal use was compared with animal use using the default starting dose of 300 mg/kg. Using the RC rat-only weight regression excluding substances with specific mechanisms of toxicity, the highest animal savings for both test methods were for substances with $2000 < LD_{50} \leq 5000$ mg/kg (1.23 [11.0%] - 3.07 [25.8%] animals) and substances with $LD_{50} > 5000$ mg/kg (3.79 [31.8%] - 4.38 [36.5%] animals). Using the NRU IC_{50} values to estimate starting doses for the ATC refined animal use by producing approximately 0.6 to 0.7 fewer animal deaths than when the default starting dose of 300 mg/kg was used.

Practical Considerations
Practical issues to consider for implementation of these cell culture test methods include the need for and availability of specialized equipment, training and expertise requirements, cost considerations, and time expenditure. Good Cell Culture Practice: ECVAM Good Cell Culture Practice Task Force Report 1 (Hartung et al. 2002) encourages the establishment of

practices and principles that will reduce uncertainty in the development and application of *in vitro* test methods.

All equipment and supplies are readily available. The NRU test methods are easily transferable to laboratories experienced with mammalian cell culture methods. Much of the training and expertise needed to perform the 3T3 and NHK NRU test methods are common to all mammalian cell culturists. Additional technical training would not be intensive since these test methods are similar in general performance to other *in vitro* mammalian cell culture assays. GLP training should be provided to technicians to ensure proper adherence to protocols and documentation procedures.

Prices for commercial testing for one substance are $1120 to $1850 for *in vitro* NRU cytotoxicity testing to determine the IC_{50} (IIVS, personal communication). It is not clear if the price of an *in vivo* test would be reduced if it were preceded by an *in vitro* cytotoxicity test to set the starting dose. Thus, use of these test methods may not reduce the overall cost of the *in vivo* rat acute oral toxicity test and might increase the cost, but their use can reduce the number of animals needed for a study. Based on cost and technical procedures associated with culture maintenance, the 3T3 cells are less expensive to use and less difficult to maintain than the NHK cells.

Peer Review

ICCVAM has considered the information in this BRD and developed draft recommendations regarding the current uses of these *in vitro* cytotoxicity test methods, and recommendations for future efforts that should be undertaken to advance the usefulness of *in vitro* methods for predicting *in vivo* acute oral toxicity. These draft recommendations are provided in a separate document. As part of the ICCVAM test method evaluation process, an independent international peer review panel will be convened to carry out an independent peer review of the 3T3 and NHK NRU test methods and to comment on the extent that the ICCVAM recommendations are supported by the information and data provided in the BRD. ICCVAM will consider the peer review panel report and public comments, and develop final test method recommendations that will be forwarded to U.S. Federal agencies for their

consideration, and where appropriate, incorporation into applicable test guidelines, regulations, and policies.

ICCVAM has also drafted test method performance standards for *in vitro* acute toxicity test methods as a separate document. These proposed standards used the NICEATM/ECVAM validation study results as performance criteria for the future use of *in vitro* test methods to determine starting doses for acute systemic toxicity testing. The test method performance standards may be revised if other methods with better predictability are adequately validated.

33
34
35
36
37
38
39
40
41
42
43
44
45
46
47   ***[This Page Intentionally Left Blank]***
48

48   **1.0   INTRODUCTION AND RATIONALE FOR THE USE OF *IN VITRO***

49   **NEUTRAL RED UPTAKE CYTOTOXICITY TEST METHODS TO**

50   **PREDICT STARTING DOSES FOR *IN VIVO* ACUTE ORAL SYSTEMIC**

51   **TOXICITY TESTING**

52

53   Poisoning is a more serious public health problem than is generally recognized. The Institute

54   of Medicine estimates that more than 4 million poisoning episodes occur annually in the

55   United States (Institute of Medicine 2004). In 2001, 30,800 deaths placed poisoning as the

56   second leading cause of injury-related death behind automobile accidents (42,433 deaths)

57   (Institute of Medicine 2004). The hazard potential for poisoning in humans is assessed by

58   acute oral toxicity testing in rodents, which is a regulatory requirement for many substances

59   and products. However, ethical and societal demands call for decreasing the numbers of

60   animals used for such studies.

61

62   *In vitro* cytotoxicity methods have been evaluated as a means to reduce and refine[1] the use of

63   animals in toxicity testing. In 1983, an international effort called the Multicentre Evaluation

64   of *In Vitro* Cytotoxicity (MEIC) was initiated to evaluate the relationship of *in vitro*

65   cytotoxicity to acute *in vivo* toxicity. Tests of 50 substances in 61 *in vitro* assays identified a

66   battery of three human cell line assays that were correlated to human lethal blood

67   concentrations. The Registry of Cytotoxicity (RC), a database that currently consists of *in*

68   *vivo* acute toxicity data from rats and mice and *in vitro* cytotoxicity data from multiple cell

69   lines for 347 substances, was published in 1998 (Halle 1998). A regression model

70   constructed from these data was proposed by ZEBET, the German National Center for the

71   Documentation and Evaluation of Alternative Methods to Animal Experiments, as a method

72   to reduce animal use by identifying the most appropriate starting doses for acute oral

73   systemic toxicity tests (Halle 1998; Spielmann et al. 1999). In October, 2000, these

74   initiatives, a European Center for the Validation of Alternative Methods (ECVAM) testing

75   strategy (Seibert et al. 1996), and other initiatives (ICCVAM 2001a [see Section 2.4, pg. 24])

---

[1] A reduction alternative is a new or modified test method that reduces the number of animals required. A refinement alternative is a new or modified test method that refines procedures to lessen or eliminate pain or distress in animals or enhances animal well-being (ICCVAM 2003).

76   to use *in vitro* cytotoxicity test methods to reduce animal use in acute toxicity testing were

77   evaluated by the International Workshop on *In Vitro* Methods for Assessing Acute Systemic

78   Toxicity (hereafter referred to as "Workshop 2000"; ICCVAM 2001a). This workshop was

79   organized by the Interagency Coordinating Committee on the Validation of Alternative

80   Methods (ICCVAM) and The National Toxicology Program (NTP) Interagency Center for

81   the Evaluation of Alternative Toxicological Methods (NICEATM).

82

83   ICCVAM recommended (ICCVAM 2001a) further evaluation of the use of *in vitro*

84   cytotoxicity data as one of the factors used to estimate starting doses for *in vivo* acute

85   lethality studies based on preliminary information that this approach could reduce the number

86   of animals used in *in vivo* studies (i.e., reduction), minimize the number of animals that

87   receive lethal doses (i.e., refinement), and avoid underestimating hazard. ICCVAM

88   concurred with the Workshop recommendation that near-term validation studies should focus

89   on two standard basal cytotoxicity assays: one using a human cell system and one using a

90   rodent cell system. Since the murine BALB/c 3T3 cytotoxicity assay had been evaluated for

91   only a limited number of chemical classes, there is merit in determining its usefulness with a

92   broader array of chemical classes. A background of historical data for *in vitro* cytotoxicity

93   testing using 3T3 cells is available through other publications (e.g., Balls et al. 1995;

94   Brantom et al. 1997; Gettings et al. 1991, 1994a, 1994b; Spielmann et al. 1991, 1993, 1996).

95   Human cell lines should also be considered since one of the aims of toxicity testing is to

96   make predictions of potential toxicity in humans (ICCVAM 2001a – *ICCVAM*

97   *Recommendations*). Historical data for *in vitro* cytotoxicity testing using normal human

98   keratinocyte (NHK) cells is also available through other publications (e.g., Gettings et al.

99   1996; Harbell et al. 1997; Sina et al. 1995; Willshaw et al. 1994).

100

101   NICEATM, in partnership with ECVAM, designed a multi-laboratory validation study to

102   evaluate animal reduction when using two mammalian cell types for *in vitro* basal

103   cytotoxicity test methods with a neutral red uptake (NRU) cell viability endpoint to predict

104   starting doses (i.e., estimated rat LD_{50} values where LD_{50} is median lethal dose) for acute oral

105   systemic toxicity test methods. The objectives for the NICEATM/ECVAM validation study

106   were to:

107	• further standardize and optimize two *in vitro* NRU cytotoxicity protocols using
108	mouse fibroblast (BALB/c) 3T3 cells and normal human epidermal
109	keratinocytes (NHK) in order to maximize intra- and inter-laboratory
110	reproducibility
111	• refine the prediction model drawn from the ZEBET approach
112	• assess the accuracy of the two standardized *in vitro* basal cytotoxicity test
113	methods for estimating rodent oral LD_{50} values across the five Globally
114	Harmonized System of Classification and Labelling of Chemicals (GHS; United
115	Nations [UN] 2005) categories of acute oral toxicity as well as unclassified
116	toxicities and estimating human lethal serum concentrations
117	• estimate the reduction and refinement in animal use achievable from using *in*
118	*vitro* basal cytotoxicity assays as one of the factors of the weight-of-evidence to
119	identify starting doses for specific *in vivo* acute toxicity tests
120	• generate high quality *in vivo* lethality and *in vitro* cytotoxicity databases that
121	can be used to support the investigation of other *in vitro* test methods necessary
122	to improve the prediction of acute systemic toxicity
123	
124	**Section 1** of this background review document (BRD) summarizes the background
125	information on the use of *in vitro* cytotoxicity test methods for predicting starting doses for
126	acute systemic toxicity assays. It includes an overview of the correlation between *in vitro*
127	cytotoxicity and acute lethality, the regulatory requirements for acute systemic toxicity
128	testing, the purpose of using *in vitro* NRU assays to predict starting doses for *in vivo* acute
129	oral systemic toxicity assays, the scientific basis of the approach, and the intended uses and
130	applicability of this approach. **Section 2** describes the protocols used to evaluate the NRU
131	assays using 3T3 and NHK cells. **Section 3** describes the selection of the reference
132	substances tested in the current validation study. **Section 4** describes the derivation of
133	reference *in vivo* rat and mouse LD_{50} values for the substances used to assess the
134	performance of the *in vitro* NRU cytotoxicity test methods (hereafter referred to as "[3T3
135	and/or NHK] NRU test methods"). **Section 5** provides the 3T3 and NHK NRU data obtained
136	during the validation study. **Section 6** refines the ZEBET approach and provides an
137	assessment of the accuracy of the NHK and 3T3 assays for predicting acute systemic

138   toxicity. **Section 7** describes the assessment of the reproducibility of the assays. **Section 8**

139   summarizes the quality of the 3T3 and NHK NRU data. **Section 9** summarizes relevant data

140   from other studies using *in vitro* cytotoxicity test methods**. Section 10** discusses computer

141   simulation modeling methods and results from the use of the 3T3 and NHK NRU test

142   methods to reduce and refine animal use in acute systemic toxicity assays. **Section 11**

143   discusses resource needs (e.g., equipment, training, time, cost) to implement these *in vitro*

144   test methods. **Section 12** provides the references and **Section 13** provides a glossary of terms

145   used in this BRD. The appendices provide supporting information for the aforementioned

146   sections.

147

148   **1.1   Background and Rationale for the Use of *In Vitro* Cytotoxicity Assays to**

149   **Predict Starting Doses for *In Vivo* Acute Oral Systemic Toxicity Tests**

150

151   Workshop 2000 was jointly sponsored by the U.S. National Institute of Environmental

152   Health Sciences (NIEHS), the NTP, and the U.S. Environmental Protection Agency (EPA).

153   During this workshop, participants reviewed the status of several major international *in vitro*

154   initiatives directed toward using *in vitro* test methods to reduce the use of laboratory animals

155   for acute toxicity testing (ICCVAM 2001a). **Sections 1.1.1** to **1.1.3** review three major

156   initiatives evaluated by Workshop 2000 participants. **Section 1.1.4** provides information on

157   the development of the NICEATM/ECVAM *In Vitro* NRU Cytotoxicity Validation Study.

158

159   1.1.1   The MEIC Program

160   The Scandinavian Society for Cell Toxicology established the MEIC program in 1983 to

161   investigate the relevance of *in vitro* test results for predicting the acute toxicity of substances

162   in humans (Bondesson et al. 1989). The program was an open study that invited interested

163   laboratories worldwide to participate in testing 50 reference substances in their particular *in*

164   *vitro* cytotoxicity assays. Although participating laboratories were requested to buy high

165   purity chemicals, no effort was made to assure that all laboratories tested substances of the

166   same purity or even purchased them from the same supplier (Clemedson et al. 1996a).

167   Minimal methodological directives were provided to maximize protocol diversity among the

168   96 participating laboratories.

169   The reference substances were selected to represent different classes of chemicals with good

170   data on acute toxicity (i.e., lethal doses, kinetics, and blood/serum concentrations [LC] in

171   humans and the oral dose producing lethality in 50% of the animals [oral LD_{50} values] in rats

172   and mice) to serve as reference values for the *in vitro* tests (Bondesson et al. 1989). The

173   MEIC management team collected human data from clinical and forensic toxicology

174   handbooks and case reports from human poisonings (Ekwall et al. 1998a). The data were

175   presented and analyzed in a series of 50 MEIC Monographs. Rat and mouse oral LD_{50} data

176   were collected from the Registry of Toxic Effects for Chemical Substances (RTECS[®]) from

177   the U.S. National Institute for Occupational Safety and Health ([NIOSH]; now licensed to

178   MDL Information Systems, Inc.).

179

180   The 50 reference substances were tested in 61 different *in vitro* assays (Ekwall et al. 1998b).

181   The measurement of interest was the concentration producing 50% inhibition of the endpoint

182   measured (i.e., IC_{50}, the concentration that produces 50% inhibition of the endpoint

183   measured). Of the 20 assays that used human-derived cells, 18 used cell lines and two used

184   primary cell cultures. Twenty-one assays used cells of animal origin (12 cell lines and nine

185   primary cell cultures). Eighteen assays were ecotoxicological tests and two were cell-free

186   test systems. The majority of the assays measured cell viability and/or cell growth.

187

188   The predictability of *in vivo* acute toxicity from the *in vitro* IC_{50} data was assessed against

189   human LC values compiled from three different data sets: clinically measured acute lethal

190   serum concentrations, acute lethal blood concentrations measured post-mortem, and peak LC

191   values derived from approximate LC_{50} curves over time after exposure (Ekwall et al. 2000).

192   A partial least squares (PLS) analysis indicated that the 61 assays predicted the three sets of

193   lethal blood concentrations well (R^2 = 0.77, 0.76 and 0.83, Q^2 = 0.74, 0.72, and 0.81,

194   respectively, where R^2 is the determination coefficient and Q^2 is the predicted variance

195   according to cross-validation in the PLS model used). The prediction of human lethal doses

196   by rat and mouse oral LD_{50} values with a two component PLS model was less accurate (R^2 =

197   0.65, Q^2 = 0.64) than the *in vitro* predictions of lethal blood concentrations.

198

199   The exposure duration for the *in vitro* assays was most often 24 hours, but ranged from 5

200   minutes to 6 weeks (Clemedson et al. 1996). Results suggested that basal (general)

201   cytotoxicity can be assessed using a variety of mammalian cell lines and almost any

202   growth/viability endpoint.

203

204   The MEIC analysis showed that the most predictive *in vitro* assays generally used human cell

205   lines (Ekwall et al. 1998b). The MEIC study yielded a battery of *in vitro* assays with good

206   performance for predicting acute lethality in humans (Ekwall et al. 2000). The MEIC team

207   concluded that improvements were necessary for *in vitro* tests to be used as complete

208   replacements for acute animal tests. To adjust for toxicity produced by mechanisms other

209   than basal cytotoxicity, the evaluation-guided development of new *in vitro* tests (EDIT) was

210   proposed to address targeted development of *in vitro* methods for other endpoints including

211   biokinetics (gut absorption, distribution, clearance), biotransformation, and target organ

212   toxicity (Clemedson et al. 2002).

213

214   1.1.2   The RC

215   The RC is a database of acute oral LD_{50} values for rats and mice obtained from RTECS® and

216   IC_{50} values from *in vitro* cytotoxicity assays using multiple cell lines and cytotoxicity

217   endpoints for substances with known molecular weights (Halle 1998). The main purpose for

218   compiling the RC was to evaluate, with a large amount of data from substances with a wide

219   range of systemic oral toxicities, whether basal cytotoxicity (averaged over various cells, cell

220   lines, and/or toxicity endpoints) is a sufficiently accurate predictor of acute systemic toxicity.

221   The RC currently contains data for 347 substances (Halle 1998) and efforts are underway to

222   increase the number of substances to 500 (ICCVAM 2001a). To date, mixtures of chemicals

223   have not been evaluated.

224

225   The RC includes published data for substances that met the following criteria for cytotoxicity

226   data (Halle 1998):

227   •   at least two different IC_{50} values were available, either from different cell types,

228   different cell lines, or different cytotoxicity endpoints

229   •   mammalian cells, with the exception of hepatocytes were used

230   • substance exposure duration was at least 16 hours, with no upper limit

231

232 The following cytotoxicity endpoints were accepted:

233   • cell proliferation: cell number, cell protein, DNA content, DNA synthesis, [3]H-
234   thymidine intake, colony formation

235   • cell viability and metabolic indicators: metabolic inhibition test (MIT-24),
236   mitochondrial reduction of tetrazolium salts into an insoluble (3-(4,5-
237   dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide [MTT]) or soluble (2,3-
238   bis(2-methoxy-4-nitro-5- sulfophenyl)-2H-tetrazolium-5-carboxanilide [XTT])
239   dye

240   • cell viability/membrane indicators: NRU, Trypan blue exclusion, cell
241   attachment, cell detachment

242   • differentiation indicators, such as functional and morphological indicators
243   within cell clusters, and/or intracellular morphology

244

245 IC_{50} values for 347 substances were obtained from 157 original publications (Halle 1998).

246 The 1,912 IC_{50} values, two to 32 per substance, were averaged using geometric means to

247 produce one IC_{50x} value for each substance.

248

249 For the RC *in vivo* data, LD_{50} values published in RTECS® were used. For the first 117

250 substances, designated as the training data set (RC-I), LD_{50} values were not revised when

251 subsequent issues of RTECS® reported different LD_{50} values. For the most recent 230

252 substances, designated as the verification set (RC-II), the LD_{50} values were taken only from

253 the 1983/84 RTECS® publication. Whenever obtainable, oral LD_{50} data from rats were used

254 (282 values). If rat data were unavailable, LD_{50} data from mice were used (65 values).

255 Combining rat and mouse data in the regression was deemed to be justified when separate

256 regressions for the mouse and rat LD_{50} data against the IC_{50x}[2] data did not result in significant

257 differences between the slopes and intercepts of the rat and mouse regressions (Halle 1998).

258

_____

[2] IC_{50x} is the geometric mean of multiple IC_{50} values collected for each substance in the RC database.

259   To obtain a model for the prediction of LD_{50} values from IC_{50} values, Halle (1998) calculated

260   a linear regression from pairs of the log-transformed IC_{50x} values (in mM) and log

261   transformed rodent oral LD_{50} values (in mmol/kg) (see **Figure 1-1**). The regression, referred

262   to here as the "RC millimole regression," has the following formula:

263

264   **log LD_{50} (mmol/kg) = 0.435 x log IC_{50x} (mM) + 0.625**

265

266   Presumably, the substance units were expressed in moles because moles are the units that

267   produce biological activity and, hence, are expected to produce the best fitting regression.

268   All of the substance data were obtained for single chemicals; chemical mixtures were not

269   included in the database and therefore were not available for determining the regression

270   formula.

271

272   To identify an acceptability range for practical use and research purposes, the acceptable

273   prediction interval for the LD_{50} was empirically defined as approximately one-half order of

274   magnitude on either side of the best-fit linear regression (i.e., ± log 5, or ± 0.699) (Halle

275   1998). This interval was based on eight linear regressions calculated for *in vitro* cytotoxicity

276   data, using various endpoints and mammalian cells, and *in vivo* rat, mouse, or rat and mouse

277   LD_{50} data from five publications. It approximates the predicted LD_{50} range for the eight

278   regressions across about eight orders of magnitude of IC_{50} values. Seventy-four percent of

279   the RC substances fall within the prediction interval.

280   **Figure 1-1**   **RC Millimole Regression Between *In Vitro* Cytotoxicity**

281   **(IC_{50x}) and Rat and Mouse Acute Oral LD_{50} Values for 347 Chemicals**

282
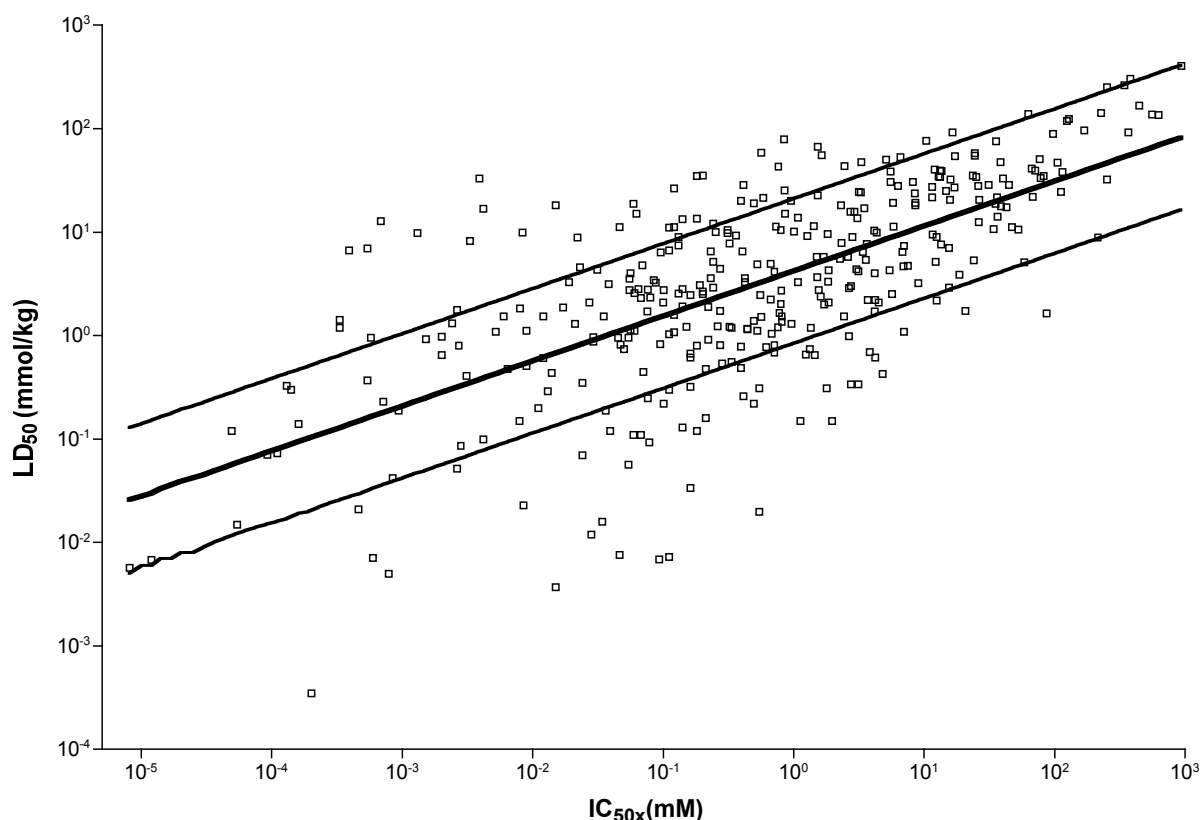
283   The heavy line shows the fit of the data to a linear regression model, $\log (LD_{50}) = 0.435 \times \log (IC_{50x})$

284    $+ 0.625$; $r=0.67$. IC_{50x} values are the geometric means of multiple endpoints and cell types. The

285   thinner lines show the empirical prediction interval ($\pm \log 5$, or ± 0.699) that is based on the

286   anticipated precision for the prediction of LD_{50} values from cytotoxicity data (Halle 1998).

287

288

289   1.1.3   <u>The ZEBET Initiative to Reduce Animal Use</u>

290   The concept that the predicted LD_{50} value could be used as a starting dose for acute oral

291   toxicity testing to reduce the number of animals was first discussed at an ECVAM workshop

292   (Seibert et al. 1996) as it related to the, then new, sequential dosing methods such as the

293   Acute Toxic Class method (ATC; OECD draft TG 423 [ICCVAM 2001a]) and the Up-and-

294   Down Procedure (UDP; OECD draft TG 425 [ICCVAM 2001a]). In these tests, for which

295   the OECD guidelines have now been finalized, the number of animals needed depends upon

296   the choice of the starting dose, since the number of consecutive dosing steps (and thus the

297   number of animals used) is reduced as the starting dose more closely approximates the true

298   toxicity class (ATC or Fixed Dose Procedure [FDP]), or the true LD_{50} (UDP). The ZEBET

299   approach involves using an IC_{50} value from an *in vitro* basal cytotoxicity test to predict an

300   LD_{50} close to the true LD_{50}. The IC_{50} is used in the RC millimole regression to predict an

301   LD_{50} value for use as a starting dose for the ATC or UDP (Spielmann et al. 1999). The use

302   of *in vitro* cytotoxicity assays to predict a starting dose equivalent to the LD_{50} may reduce

303   animal use in the UDP by 25-40%, depending upon the slope of the curve and the stopping

304   rule applied (Spielmann et al. 1999; ICCVAM 2001a).

305

306   1.1.4   The NICEATM/ECVAM *In Vitro* NRU Cytotoxicity Validation Study

307   Workshop 2000 participants concluded that none of the *in vitro* models reviewed had been

308   formally evaluated for reliability and relevance, and their usefulness and limitations for

309   generating information for acute toxicity testing had not been assessed. However, the

310   approach proposed by ZEBET (Halle 1998; Spielmann et al. 1999) was recommended for

311   rapid adoption so that data could be generated to establish its usefulness with a large number

312   of substances (ICCVAM 2001a). To assist in the adoption and implementation of the

313   ZEBET approach, several workshop participants wrote *Guidance Document on Using In*

314   *Vitro Data to Estimate In Vivo Starting Doses for Acute Toxicity* (hereafter referred to as

315   *Guidance Document*; ICCVAM 2001b).

316

317   The *Guidance Document* recommended testing 10 to 20 reference substances of high purity

318   from the RC in a candidate *in vitro* basal cytotoxicity assay to be used for predicting starting

319   doses for acute oral lethality tests (ICCVAM 2001b). The substances were to cover a wide

320   range of toxicity and fit the RC prediction model (i.e., the linear regression line) as closely as

321   possible. The assays recommended and provided as examples are NRU assays using 3T3

322   and NHK cells. The IC_{50} results for the selected substances would be used to calculate a new

323   regression line with the LD_{50} values used by the RC. If the resulting regression were parallel

324   to the RC millimole regression and within the $\pm \log 5$ (i.e., ± 0.699) prediction interval for

325   the RC, the *Guidance Document* recommended using the cytotoxicity assay to predict

326   starting doses for LD_{50} assays. If the regression from the assay did not meet these criteria,

327   then the *Guidance Document* advised either (a) adjusting the slope or (b) using the NRU

328   protocols offered in the *Guidance Document* (considered the most efficient approach).

329   To further characterize the usefulness of the 3T3 and NHK NRU test methods as predictors

330   of starting doses for acute oral systemic toxicity assays, NICEATM and ECVAM designed

331   an independent[3] multi-laboratory validation study to evaluate the performance of these *in*

332   *vitro* test methods. The inclusion of human cells in the NICEATM/ECVAM validation study

333   implements a Workshop 2000 recommendation to evaluate whether cytotoxicity in human or

334   rodent cells best predicts human acute toxicity. ECVAM's development of a prediction

335   model for human acute toxicity using data collected in the NICEATM/ECVAM validation

336   study will be addressed elsewhere.

337

338   *Study Design*

339   The planning phases of the NICEATM/ECVAM validation study included the selection of

340   reference substances for testing, which is described in **Section 3,** and the identification of

341   reference LD_{50} values for the reference substances, which is described in **Section 4**. The

342   NRU testing proceeded in several phases (See **Figure 1-2**) so that the Study Management

343   Team (SMT) could evaluate the reproducibility of results after each phase and refine the

344   protocols, if necessary, before proceeding to the next phase. The NRU data collected during

345   the laboratory phase were used to evaluate, and in some cases, develop, linear regression

346   formulas for the prediction of LD_{50} values by IC_{50} values (see **Section 6**). Computer

347   simulation modeling of acute oral toxicity test outcomes was then performed to determine

348   animal savings using the NRU-predicted starting doses compared with the default starting

349   dose (see **Section 10**). Study management and study participant information is provided in

350   **Appendix A.**

351

---

[3] "Independent" is used here to indicate that neither NICEATM nor ECVAM neither developed nor had
monetary interest in the test methods.

351   **Figure 1-2   NICEATM/ECVAM Validation Study Phases**

352   **Phase Ia: Laboratory Evaluation**

353   Development of a positive control database for each laboratory
354   •   Perform at least 10 replicate NRU tests of the positive control substance (sodium laurel
355   sulfate [SLS]) with each cell type.
356   •   Calculate mean $IC_{50} \pm 2$ SD for each cell type for each lab.
357   •   Establish acceptance criteria for positive control performance in future assays.
358
359   ⇓

360   **Phase Ib: Laboratory Evaluation**

361   Limited substance testing to demonstrate the reliability of the protocol
362   •   Each laboratory tests the same three coded substances of varying toxicities three times
363   with each cell type.
364   •   Refine protocols and repeat, if necessary, until acceptable intra/interlaboratory
365   reproducibility is achieved.
366
367   ⇓

368
369   **Phase II: Laboratory Qualification**

370   Evaluation of protocol refinements
371   •   Each laboratory tests nine coded substances covering the range of GHS toxicity
372   categories, with three replicate tests/substance for each test method.
373   •   Assure that corrective actions taken in Phase I have achieved the desired results.
374   •   Further refine protocols and re-test, if necessary, to achieve acceptable reliability.
375   •   Finalize protocols for Phase III.
376
377   ⇓

378
379   **Phase III: Laboratory Testing Phase**

380   Test of optimized protocols
381   •   Each laboratory tests 60 coded substances three times using the final protocol for each
382   test method.
383
384

385

386

386 **1.2   Regulatory Rationale and Applicability for the Use of *In Vitro* Cytotoxicity**

387   **Test Methods to Predict Starting Doses for Acute Oral Systemic Toxicity**

388   **Testing**

389

390 1.2.1   Current Regulatory Testing Requirements for Acute Systemic Toxicity

391 The major regulatory requirement for acute systemic toxicity testing is for the hazard

392 classification and labeling of products, which is intended to protect handlers and consumers

393 from toxic hazards. The LD_{50} results from acute systemic toxicity tests are used to place

394 substances in various toxicity categories that, in turn, invoke the associated hazard phrases to

395 be used on product labels. **Table 1-1** shows the current U.S. legislation requiring the use of

396 acute systemic toxicity testing for product labeling and the substances regulated. **Table 1-2**

397 shows the statutory protocol requirements and classification systems used by each U.S.

398 regulatory agency. Also included is an international guideline for labeling, the Harmonized

399 Integrated Classification System for Human Health and Environmental Hazards of Chemical

400 Substances and Mixtures, which provides guidance to regulatory agencies on the use of the

401 GHS (UN 2005) as a method for an internationally comprehensible system for hazard

402 communication (OECD 2001b).

403

404 **Table 1-1   Summary of Current U.S. Legislation for Using Acute Systemic Toxicity**

405   **Data for Product Labeling**

Legislation (Year of Initial Enactment)	U.S. Regulatory Agency	Substance
Federal Insecticide, Fungicide and Rodenticide Act (1947)	EPA	Pesticides
Federal Hazardous Substances Act (1964)	CPSC	Household products
Occupational Safety and Health Act (1970)	OSHA	Occupational materials
Federal Hazardous Material Transportation Act (1975)	DOT	Transported substances

406 Abbreviations: EPA = U.S. Environmental Protection Agency; CPSC = U.S. Consumer Product Safety
407 Commission; OSHA = U.S. Occupational Safety and Health Administration; DOT = U.S. Department of
408 Transportation.

409 Note: The U.S. Food and Drug Administration (FDA) does not require data for acute lethality testing, and in
410 fact, discourages the use of animals for such testing (FDA 1993).

411

412 **Table 1-2   Regulatory Classification Systems for Acute Oral Toxicity**

413

Regulatory Agency (Authorizing Act)	Animals	Endpoint	Classification
EPA (Federal Insecticide, Fungicide and Rodenticide Act)	Use current EPA or OECD protocol	Death[1]	I - $LD_{50} \leq 50$ mg/kg II - $50 < LD_{50} \leq 500$ mg/kg III – $500 < LD_{50} \leq 5000$ mg/kg IV - $LD_{50} > 5000$ mg/kg
CPSC (Federal Hazardous Substances Act)	White rats, 200-300 g	Death[1] within 14 days for $\geq$ half of a group of $\geq$ 10 animals	Highly toxic – $LD_{50} \leq 50$ mg/kg Toxic – 50 mg/kg $< LD_{50} < 5$ g/kg
OSHA (Occupational Safety and Health Act)	Albino rats, 200-300 g	Death[1], duration not specified.	Highly toxic - $LD_{50} \leq 50$ mg/kg Toxic – $50 < LD_{50} < 500$ mg/kg
DOT (Federal Hazardous Material Transportation Act)	Male and female young adult albino rats	Death[1] within 14 days of half the animals tested. Number of animals tested must be sufficient for statistically valid results.	Packing Group 1 - $LD_{50} \leq 5$ mg/kg Packing Group II – $5 < LD_{50} \leq 50$ mg/kg Packing Group III – $LD_{50} < 500$ mg/kg (liquid) $\quad\quad\quad LD_{50} < 200$ mg/kg (solid)
OECD Guidance for Use of GHS (2001a)	Protocol not specified	Protocol not specified	I - $LD_{50} \leq 5$ mg/kg II - $5 < LD_{50} \leq 50$ mg/kg III - $50 < LD_{50} \leq 300$ mg/kg IV - $300 < LD_{50} \leq 2000$ mg/kg V - $2000 < LD_{50} \leq 5000$ mg/kg Unclassified - $LD_{50} > 5000$ mg/kg

414 [1]*Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints*
415 *for Experimental Animals Used in Safety Evaluation* calls for humane killing of moribund animals
416 (OECD 2000). Moribund animals that are humanely euthanized are accepted as deaths.
417 Abbreviations: EPA = U.S. Environmental Protection Agency; CPSC = U.S. Consumer Product Safety
418 Commission; OECD = Organisation for Economic Co-operation and Development; OSHA = U.S.
419 Occupational Safety and Health Administration; DOT = U.S. Department of Transportation; GHS =
420 Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005)

421

422　In addition to classification and labeling, acute systemic toxicity test results may be used for:

423　　　• establishing dosing levels for repeated dose toxicity studies

424　　　• generating information on the specific organs affected

425　　　• providing information related to the mode of toxic action

426　　　• aiding in the diagnosis and treatment of toxic reactions

427　　　• providing information for comparison of toxicity and dose response among

428　　　　substances in a specific chemical or product class

429　　　• aiding in the standardization of biological products

430　　　• aiding in judging the consequences of single, high accidental exposures in the

431　　　　workplace, home, or from accidental release

432　　　• serving as a standard for evaluating alternatives to animal tests

433

434　*Test Methods for Assessing Acute Systemic Toxicity*

435　The current internationally recognized test methods for acute systemic toxicity testing are the

436　FDP (OECD 2001c), the ATC method (OECD 2001d), and UDP (OECD 2001a; EPA 2002a)

437　(see **Appendix M** for test method guidelines). Information on signs of acute toxicity and

438　target organs can be obtained using any of the three test methods. All three methods are

439　sequential tests in which the outcome of testing one or more animals at the first dose is used

440　to determine the second dose that should be tested. The FDP differs from the UDP and ATC

441　in that it involves testing more animals per dose and the primary endpoint of interest is

442　evident toxicity[4] rather than lethality. Both the FDP and the ATC method provide a range for

443　the LD_{50} for classification purposes. The UDP generally provides a point estimate of the

444　LD_{50} with a confidence interval (EPA 2002a).

445

446　Each of the test method guidelines include a limit test in which up to five (UPD and FDP) or

447　six (ATC) animals are tested at the limit, or upper bound, dose (OECD 2001a,c,d; EPA

448　2002a). The limit test can be performed using 2000 or 5000 mg/kg.

449

---

[4] *Evident toxicity* is a general term describing clear signs of toxicity following administration of test substance, such that an increase to the next highest fixed dose would result in the development of severe toxic signs and probably mortality (ICCVAM 2000).

450   1.2.2   Intended Regulatory Uses for the *In Vitro* Cytotoxicity Test Methods

451   *In vitro* cytotoxicity test methods are not recommended for the replacement of acute oral

452   toxicity tests in animals. Rather, such test methods are intended to serve as adjuncts for *in*

453   *vivo* acute systemic toxicity test methods. To select a starting dose for a test substance, the

454   current test guidelines for acute oral systemic toxicity recommend using information on

455   structurally-related substances and the results of any other toxicity tests (EPA 2002b),

456   including *in vitro* cytotoxicity results (OECD 2001a, c, d; EPA 2002a). The 3T3 and NHK

457   NRU test methods are intended to be used as part of the weight-of-evidence approach to

458   select starting doses for the UDP and ATC assays in order to reduce and refine the use of

459   animals for *in vivo* acute toxicity testing. The reduction of animals achievable with the use

460   of basal cytotoxicity as an adjunct to the UDP or ATC is provided in **Section 10**. Since the

461   estimation of the true LD_{50} is irrelevant to setting doses for measuring evident toxicity, the

462   FDP will not be considered further in this document.

463

464   **Section 10** presents analyses that characterize the extent of animal reduction and refinement

465   that may occur by using the NRU test methods to estimate the starting doses for the UDP and

466   the ATC method. Animal use and animal deaths for UDP and ATC testing is determined

467   using computer simulation techniques rather than by animal testing. The simulations of UDP

468   and ATC testing determine the number of animals used when using the default starting dose

469   and when using a starting dose determined from the NRU test methods. The number of

470   animals used with the NRU-determined starting dose is compared with the number of

471   animals used with the default starting dose to determine the reduction in animal use with the

472   NRU-determined starting dose. To characterize the extent of refinement produced by using

473   the NRU-determined starting dose, the number of animals that die with the NRU-determined

474   starting dose is compared with the number of animals that die when using the default starting

475   dose.

476

477   1.2.3   Similarities and Differences in the Endpoints of the *In Vitro* Cytotoxicity Test

478   Methods and *In Vivo* Acute Oral Toxicity Test Methods

479   The endpoint measured in the *in vitro* NRU cytotoxicity test methods is cell death (neutral

480   red [NR] is taken up only by live cells) and the major endpoint of interest is the concentration

481   at 50% inhibition of NRU (i.e., the IC_{50}). The endpoint measured in acute systemic toxicity

482   assays is usually animal death. Cell death and animal death may be similar since animals are

483   comprised of organ systems consisting of tissues, which are comprised of cells. All cells,

484   regardless of whether they are in animals or *in vitro* cell cultures, have similar cellular

485   mechanisms of energy production and utilization and maintenance of cell membrane

486   integrity.

487

488   Animal death and death of cells in culture due to toxicity are similar in that both involve

489   some type of cellular injury. For the animal, the cellular injury produces tissue and organ

490   injury to the most sensitive target organ, which may then cause the death of the whole

491   organism. Organ system failure can be due either to the death of cells in the affected organ

492   or to the loss of function of the surviving cells in the organ, which results in cell death or loss

493   of function in other organs (Gennari et al. 2004). Death of an animal is produced by major

494   organ system failure. Ultimately the cardiovascular and respiratory systems fail. Respiratory

495   depression may be due to depression of the central nervous system (CNS) rather than a direct

496   assault on the respiratory system. Other major organ system failures, such as liver and

497   kidney failure, gastrointestinal corrosion, and bone marrow depression, also produce death.

498   Cell death in a culture system involves the death of a single cell type. Cell death and animal

499   death may be produced by the same mechanisms, such as disruption of membrane structure

500   or function, inhibition of mitochondrial function, disturbance of protein turnover, disruption

501   of energy production, etc. (Gennari et al. 2004).

502

503   Animal and cell culture systems are different with respect to how a substance or toxin is

504   delivered to the cell and how it is distributed, metabolized, and excreted. After oral

505   administration, animals must absorb the toxin from the gastrointestinal tract, which involves

506   the passage of membranes. The toxin may or may not be heavily bound to serum proteins;

507   this would reduce the availability of the toxin to the target organ. The toxin may then be

508   metabolized during and/or after distribution to the target organs and then the toxin or its

509   metabolites are excreted. In a cell culture system, the only membranes that must be passed

510   are those of the target cell and cellular organelles. No absorption and distribution by other

511   cellular systems is required. Cell culture systems may or may not include serum proteins,

512   which could reduce the availability of toxin to act as its target site. The 3T3 cell culture

513   system includes serum while the NHK cell culture system does not. The 3T3 and NHK cell

514   culture systems have little to no capacity to metabolize xenobiotic compounds. Excretion

515   from the cell culture milieu cannot occur since cell culture systems have no excretory system.

516   The cultured cells are exposed to substances for the entire duration of exposure in the test

517   system.

518

519   Animal and cell culture systems may also be different with respect to the target on which a

520   toxin acts. If a toxin acts in a specialized organ system in a whole animal, it may not produce

521   a toxic effect by the same mechanism in cultured cells that are derived from tissue different

522   from the target organ. For example, a neurotoxin that acts by a neuroreceptor-mediated

523   pathway in animals, would be expected to produce toxicity by a different mechanism in 3T3

524   or NHK cells, which are derived from fibroblasts, and skin cells, respectively. Even if a

525   neurotoxin were applied to neuronal cells in culture, the cultured cells may not respond in the

526   same way as neuronal cells in a whole animal. Cultured cells may not retain the same

527   functionality as cells *in vivo*.

528

529   1.2.4   Use of *In Vitro* Cytotoxicity Test Methods in the Overall Strategy of Hazard

530   Assessment

531   In the overall strategy of hazard or safety assessment, the intended regulatory use of *in vitro*

532   test methods is to reduce and refine the use of animals in current acute systemic toxicity

533   assays (i.e., serve as adjuncts to these test methods). *In vitro* cytotoxicity test methods are

534   not intended as replacements for the *in vivo* tests. For current OECD acute systemic toxicity

535   assays (the ATC or UDP), that use sequential dosing methods, the number of animals used

536   depends on the choice of starting dose since the number of dosing steps (and animals) is

537   reduced if the starting dose is close to the true toxicity class (ATC) or to the true LD_{50} (UDP)

538   (Spielmann et al. 1999; ICCVAM 2001b).

539

540   As noted earlier, Spielmann et al. (1999) and the *Guidance Document* (ICVAM 2001b)

541   suggest that the RC millimole regression be used with *in vitro* cytotoxicity data to predict

542   starting doses for the ATC and UDP. The approach can be applied to substances with purity

543   appreciably lower than 100% as long as molecular weight and purity are known. Therefore,

544   this approach is not applicable to mixtures such as product formulations or unknown

545   substance samples.

546

547   Thus, in addition to evaluating the reduction of animal use associated with the ATC and UDP

548   when the current RC millimole regression (in millimolar units) is used to predict the starting

549   dose, this study also evaluated the reduction in animal use associated with regressions based

550   on weight units.

551

552   **1.3   Scientific Basis for the *In Vitro* NRU Test Methods**

553

554   Cytotoxicity has been defined as the adverse effects resulting from interference with

555   structures and/or processes essential for cell survival, proliferation, and/or function (Ekwall

556   1983). Ekwall (1983) described the concept of "basal cell functions" that virtually all cells

557   possess (mitochondria, plasma membrane integrity, etc.) and suggested that, for most

558   substances, toxicity is a consequence of non-specific alterations in those cellular functions,

559   which may then lead to effects on organ-specific functions and/or death of the organism.

560   These effects may involve the integrity of membranes and the cytoskeleton, cellular

561   metabolism, the synthesis and degradation or release of cellular constituents or products, ion

562   regulation, and cell division.

563

564   Ekwall (1983) and others (Grisham and Smith 1984) concluded that, since the actions of

565   substances that produce injury and death are ultimately exerted at the cellular level, *in vitro*

566   cytotoxicity assays may be useful for the prediction of acute lethal potency. Considerable

567   research has been undertaken to develop and evaluate *in vitro* tests for use as screens and as

568   potential replacements for LD_{50} tests. Good agreement between cytotoxicity *in vitro* and

569   animal lethality have been reported by numerous groups (see reviews by Phillips et al. 1990;

570   Garle et al. 1994; Guzzie 1994). However, none of the proposed *in vitro* models have been

571   evaluated in any formal studies for reliability and relevance, and their usefulness and

572   limitations for generating information to meet regulatory requirements for acute toxicity

573   testing have not been assessed.

574   1.3.1   Purpose and Mechanistic Basis of the *In Vitro* NRU Test Methods

575   There are a number of basal cytotoxicity endpoints that measure cell death and or cell

576   proliferation. The NRU test methods were chosen for the NICEATM/ECVAM validation

577   study because they were recommended in the *Guidance Document* for the purpose of

578   obtaining cytotoxicity information to predict starting doses for acute systemic toxicity assays

579   (ICCVAM 2001b). Both the 3T3 and NHK NRU test methods were reproducible in previous

580   validation studies (ICCVAM 2001b). In addition, both cell types are easily obtainable from

581   commercial sources and the *Guidance Document* provided preliminary evidence that these

582   assays could reproduce the RC millimole regression. Additionally, the assays can be

583   automated and they require no radioactivity or highly dangerous substances (see **Section 2**

584   for the protocols).

585

586   Neutral red is a weakly cationic water-soluble dye that stains living cells (Borenfreund and

587   Puerner 1985). It readily diffuses through the plasma membrane and concentrates in

588   lysosomes where it electrostatically binds to the anionic lysosomal matrix. Toxins can alter

589   the cell surface or the lysosomal membrane seeming to cause lysosomal fragility and other

590   adverse changes that gradually become irreversible. Thus, cell death and/or inhibition of cell

591   growth decreases the amount of neutral red taken up by the culture. The protocol for the

592   NRU assay using 3T3 cells was first published by Borenfreund and Puerner (1985) as a two

593   component test for toxicity screening that was standardized for a 96-well plate format. The

594   two components were (1) a morphological examination of the cells under an inverted phase

595   microscope and (2) a quantitative measurement of NRU. The morphological examination

596   was designed to identify the highest tolerated dose for the assay (i.e., the highest

597   concentration of toxicant that the cells can tolerate and that causes minimal morphological

598   changes). This concentration was comparable to the quantitative measurement of 10%

599   inhibition (i.e., NR_{90} value compared to the controls) of NRU. The NR_{90} value is the point

600   where a test compound produces a significant toxic effect. The assay was said to be a rapid,

601   reliable, inexpensive, and reproducible *in vitro* assay for screening potentially toxic agents,

602   and it was suggested that the test was a good candidate for inclusion in a battery of tests for

603   toxicity screening for the purpose of reducing the use of animals for toxicity tests.

604

605   1.3.2   Similarities and Differences in the Modes/Mechanisms of Action for the *In Vitro*
606   NRU Test Methods Compared with the Species of Interest
607   Although the ultimate species of interest for acute systemic toxicity concerns is humans,
608   labeling and hazard identification requirements are based on rodent studies. There are
609   differences between humans and rodents in terms of absorption, distribution, metabolism,
610   excretion, and the intrinsic sensitivity of target organs to xenobiotic compounds. The
611   differences are largely substance specific. *In vitro* cytotoxicity studies have also noted
612   differences in sensitivity between human cells and other mammalian cells (Clemedson et al.
613   1996).
614
615   Ekwall et al. (1998b) showed that *in vitro* cytotoxicity methods using human cell lines
616   generally predicted human toxicity better than methods using other mammalian cell types.
617   **Section 6** shows that, for the reference substances tested in this study, the 3T3 NRU test
618   method usually predicted rodent acute toxicity better than the NHK NRU test method did. A
619   human cell type, such as the NHK, may predict human toxicity better than 3T3 cells, which
620   originate in mice (this evaluation is not reported in this BRD, but will be reported elsewhere).
621
622   Besides the species differences, there are several other differences between the 3T3 and NHK
623   cells.
624   •   The 3T3 cells are an immortal line, while the NHK cells are primary cells.
625   •   They originate from different tissues; 3T3 cells are derived from embryonic
626   fibroblasts, while the NHK cells come from neonatal foreskin tissue.
627   •   NHK cells grow more slowly in culture than the 3T3 cells.
628   •   NHK cells have greater ability to metabolize xenobiotic compounds, in that they
629   exhibit some cytochrome P450 activity (Babich et al. 1991). 3T3 cells have
630   practically no ability to metabolize xenobiotic compounds (INVITTOX 1991).
631
632   1.3.3   Range of Substances Amenable to the *In Vitro* NRU Test Methods
633   The *in vitro* NRU test methods can be applied to a wide range of substances as long as the
634   substances can be dissolved in the cell culture medium or in a solvent that can be mixed with
635   culture medium. Although these test methods may to be applicable to mixtures, none were

636   evaluated in this validation study. The toxicity of substances with specific mechanisms of

637   toxicity not expected to be active in 3T3 or NHK cells (e.g., those that are neurotoxic,

638   cardiotoxic, interfere with energy utilization, or alkylate proteins and other macromolecules)

639   will likely be underpredicted by these test methods. Therefore, until a more predictive

640   approach is developed, the results from basal cytotoxicity testing with such substances may

641   not be appropriate.

642

643   Insoluble substances or those unstable or explosive in water are not compatible with the test

644   system. Volatile substances may yield acceptable results if CO_2 permeable plastic film is

645   used to seal the test plates. Testing for corrosive substances is unnecessary since there is no

646   regulatory requirement for acute systemic toxicity testing for corrosives. The toxicity of

647   substances that are highly bound to serum proteins may be underestimated by the 3T3 assay

648   since the culture medium contains 5% serum during substance exposure. The toxicity of

649   substances that specifically affect lysosomes may be overestimated since they may affect

650   NRU. Red substances that absorb light in the optical density range of NR may interfere with

651   the test if they remain inside the cell in sufficient amounts after washing and are soluble in

652   the NR solvent.

653

63 **2.0   TEST METHOD PROTOCOL COMPONENTS OF THE 3T3 AND NHK *IN***

64 **_VITRO_ NRU TEST METHODS**

65

66 The *Guidance Document* (ICCVAM 2001b) recommended that the following conditions be

67 incorporated into any *in vitro* cytotoxicity protocol used to predict *in vivo* acute lethality:

68 - use a cell line (or primary cells) that divides rapidly

69 - use an initial seeding density that allows rapid growth throughout the exposure

70   period

71 - apply reference substances only on cells in the exponential phase of growth

72 - use a reference substance exposure period at least the duration of one cell cycle

73 - use appropriate positive and vehicle control substances for which cytotoxicity,

74   or lack of cytotoxicity, has been well characterized by the performing laboratory

75 - use solvents only at levels previously shown not to cause cytotoxicity to the cell

76   system over the entire period of the assay

77 - use a well established measurement endpoint that has good interlaboratory

78   reproducibility

79 - use tests compatible with 96-well plates and apparatus (i.e., spectrophotometers)

80   that allow a quick and precise measurement of the endpoint

81 - use a progression factor in the concentration-response experiment that yields

82   graded effects between no effect and total cytotoxicity

83

84 **Section 2.1** provides descriptions of the protocol applications to the NICEATM/ECVAM *In*

85 *Vitro* Cytotoxicity Validation Study. **Section 2.2** provides details for performing the 3T3 and

86 NHK NRU test methods and explains the rationale for various test method components. The

87 basis for the selection of these *in vitro* cytotoxicity test methods is given in **Section 2.3** and

88 proprietary aspects associated with this study are described in **Section 2.4**. **Section 2.5**

89 discusses the basis for replicate and repeat tests. **Section 2.6** details the modifications and

90 revisions made throughout all phases leading to the development of the final protocol used in

91 Phase III of this validation study. **Section 2.7** shows the differences between the test

92 methods used in this study and the test methods outlined in the *Guidance Document*.

93   **Sections 2.8** and **2.9** provide details on the solubility protocol for the reference substances

94   used in to validate the two *in vitro* NRU cytotoxicity test methods.

95

96   These test method protocols were provided to the three cytotoxicity testing laboratories that

97   participated in the NICEATM/ECVAM study (see **Section 5.6.3** for additional laboratory

98   information):

99   • ECBC: The U.S. Army Edgewood Chemical Biological Center

100   • FAL: Fund for the Replacement of Animals in Medical Experiments (FRAME)

101   Alternatives Laboratory

102   • IIVS: Institute for *In Vitro* Sciences

103

104   A fourth laboratory was used (BioReliance Corporation, Rockville, MD) to procure and

105   distribute the coded reference substances and to perform solubility tests on all validation

106   study reference substances prior to distribution to the cytotoxicity testing laboratories.

107

108   **2.1   Overview of the 3T3 and NHK NRU Test Methods**

109

110   The authors of the *Guidance Document* (ICCVAM 2001b) developed and presented a

111   proposed 3T3 NRU protocol for use in a validation study based on the BALB/c 3T3

112   Cytotoxicity Test, INVITTOX Protocol No. 46 (available at the FRAME-sponsored

113   INVITTOX database [http://embryo.ib.amwaw.edu.pl/invittox/]) which in turn was based on

114   the Borenfreund and Puerner (1985) protocol, as elaborated on in Spielmann et al. (1991) and

115   Spielmann et al. (1996).

116

117   The *Guidance Document* protocol also included revisions based on experience with a

118   modification of another test, the 3T3 NRU Phototoxicity Test, INVITTOX Protocol No. 78,

119   also available at the FRAME database. The Registry of Cytotoxicity (RC) regression for

120   prediction of acute oral systemic rodent (rat and mouse) toxicity (Halle 1998; Spielmann et

121   al. 1999) was included as the prediction model (see **Section 1.1.2**). The RC is a database of

122   acute oral LD_{50} values for rats and mice obtained from RTECS[®] and IC_{50} values from *in vitro*

123   cytotoxicity assays using multiple cell lines and cytotoxicity endpoints for chemicals with

124   known molecular weights.

125

126   The NHK NRU test method protocol in the *Guidance Document* was based on a NRU test

127   method by Borenfreund and Puerner (1984) using human epidermal keratinocytes (Heimann

128   and Rice 1983) and was obtained from IIVS. Formulations for the media and solutions and

129   general NHK cell culture techniques correspond to Clonetics® products from the CAMBREX

130   Corporation. The authors of the *Guidance Document* expanded the IIVS protocol by adding

131   details on equipment, media and reagent components, and experimental procedure.

132

133   The test method protocol components for the *in vitro* NRU cytotoxicity test methods used in

134   the NICEATM/ECVAM study are very similar for both the 3T3 and the NHK cells (see

135   **Figure 2-1**). The following procedures are common to both cell types:

136   •   preparation of reference substances and positive control

137   •   cell culture environmental conditions

138   •   determination of test substance solubility

139   •   96-well plate configuration for testing samples

140   •   range finder and definitive tests (48-hour exposure to the reference substance)

141   •   microscopic evaluation of cell cultures for toxicity

142   •   measurement of NRU

143   •   data analysis

144

145   The main differences in the test methods are:

146   •   the conditions of propagation of the cells in culture

147   •   the cell growth medium components

148   •   the application of reference substances to the 96-well plate (i.e., different

149   volumes of reference substance solution)

150

151   The nature of the NRU response is described in **Section 1.3.1**. **Figure 2-1** provides an

152   overview to the major steps for performance of the *in vitro* NRU cytotoxicity test methods.

153   **Figure 2-1   Major Steps for Performance of the NRU Test Methods in the**

154   ***In Vitro* Cytotoxicity Validation Study**

155

156   (1) 3T3 cells or NHK cells are seeded into 96-well plates to form a sub-confluent

157   monolayer (24 hours for 3T3 cells, 48-72 hours for NHK cells)

158   ⇓

159   (2) Culture medium is removed (for 3T3 cells only)

160   ⇓

161   (3) Reference substances in treatment medium are added to the cells; cells are

162   exposed for 48 hours to the reference substance over a range of eight (8)

163   concentrations

164   ⇓

165   (4) Cells are evaluated microscopically for toxicity based on morphological

166   alterations

167   ⇓

168   (5) Treatment medium is removed; cells are washed once with Dulbecco's Phosphate

169   Buffered Saline (D-PBS); Neutral Red (NR) dye medium is added (3T3 cells: 25

170   μg/mL NR dye; NHK cells: 33 μg/mL NR dye); plates are incubated for 3 hours

171   ⇓

172   (6) NR medium is discarded; cells are washed once with D-PBS; NR desorbing

173   fixative is added to the plates

174   ⇓

175   (7) Plates are shaken for 20 minutes

176   ⇓

177   (8) NR absorption is measured at optical density (OD) 540 ± 10 nm

178   ⇓

179   (9) NRU is calculated as the % of control values to define IC_{20}, IC_{50}, and IC_{80} reference

180   substance concentrations (μg/mL)[1]

181

---

[1] IC_{50} values are used for estimating the LD50 value of a reference substance. The IC_{20} and IC_{80} values were collected (as per request in the validation study's Statement of Work [SOW]) for possible use in estimating human lethal concentrations in blood.

181   2.1.1   <u>The 3T3 NRU Test Method</u>

182   *Initiating and Subculturing of 3T3 Cells*

183   (CCL-163, 3T3 BALB/c mouse fibroblast, clone 31, American Type Culture Collection

184   [ATCC], Manassas, VA, USA)

185

186   Cryopreserved 3T3 cells are thawed, resuspended in a routine culture medium containing

187   Dulbecco's Modification of Eagle's Medium (DMEM) supplemented with non heat-

188   inactivated 10% newborn calf serum (NCS), transferred into tissue culture flasks (25 or 75 -

189   80 cm^2), and incubated at 37°C ± 1°C, 90% ± 5% humidity, and 5.0% ± 1% CO_2/air. When

190   cells reach 50 – 80% confluency (as estimated from a visual inspection of cell density), they

191   are removed from the flask by trypsinization. A single-cell suspension is added to new flasks

192   for propagation and the cells are passaged/subcultured at least two times before seeding into

193   96-well plates for test assays. Subsequent passages may be maintained in culture for

194   approximately two months (~18 passages) and used in NRU test methods. A new frozen

195   ampule is thawed when needed and the above procedures are repeated. The protocols

196   provide cell culture density guidelines for subculturing the cells and each laboratory

197   determines the final seeding densities to achieve appropriate growth.

198

199   *Preparation of Cells for 96-well Plate Assays*

200   After achieving appropriate subculturing of cells, 100 µL of the cell suspension (2.0 –

201   3.0x10^3 cells/well) are placed in the appropriate wells and 100 µL of cell-free culture

202   medium are dispensed into the peripheral wells (blanks). One plate per reference substance

203   is prepared. The cells are incubated for 24 ± 2 hours and checked to be sure that

204   approximately a half-confluent monolayer is attained at the time of reference substance

205   application.

206

207   *Reference Substance Application*

208   After the appropriate incubation period, medium is removed and 50 µL of the routine culture

209   medium with 10% NCS are added to each well. Then, 50 µL treatment medium containing

210   the appropriate reference substance concentrations are added for a final concentration of 5%

211   NCS. The cells are incubated for 48 ± 0.5 hours. At the end of the incubation period, the

212 cells are microscopically evaluated for changes in morphology and their appearance is

213 documented (as per Visual Observation Codes in the protocol) prior to measurement of the

214 NRU of the cells.

215

216 2.1.2   The NHK NRU Test Method

217 *Initiating and Subculturing of NHK Cells*

218 (pooled primary neonatal foreskin cells, Clonetics[®] # CC-2507, lot # 1F0490N, CAMBREX

219 Bio Science Walkersville, Inc., Walkersville, MD, USA)

220

221 Cryopreserved cells are thawed, resuspended in keratinocyte complete growth medium,

222 transferred into tissue culture flasks (25 cm^2 without fibronectin-collagen coating), and

223 incubated at 37°C ± 1°C, 90% ± 5% humidity, and 5.0% ± 1% CO_2/air. When cells reach 50

224 – 80% confluency (as estimated from a visual inspection of cell density), they are removed

225 from the flask by trypsinization and prepared for subculturing into the 96-well plates.

226 Keratinocytes are not subcultured beyond the second passage. Additional frozen ampule(s)

227 are thawed as needed. The protocols provide cell culture density guidelines for establishing

228 the cells out of cryopreservation and each laboratory determines the final seeding densities to

229 achieve appropriate growth.

230

231 *Preparation of Cells for 96-well Plate Assays*

232 After appropriate subculturing of cells is achieved, 125 µL of the cell suspension (2.0 –

233 2.5x10^3 cells/well) are placed in the appropriate wells and 125 µL of cell-free culture

234 medium are dispensed into the peripheral wells (blanks). One plate per reference substance

235 is prepared. The cells are incubated for ~ 48 - 72 hours and checked to be sure that a

236 monolayer of 20+% confluency (e.g., 20 – 50% confluency) is attained at the time of

237 reference substance application.

238

239 *Reference Substance Application*

240 After the appropriate incubation period, 125 µL of the culture medium containing the

241 appropriate reference substance concentrations are added to the test wells (the existing 125

242 µL of culture medium is not removed). The cells are incubated for 48 ± 0.5 hours. At the

243 end of the incubation period, the cells are microscopically evaluated for changes in

244 morphology and their appearance is documented (as per Visual Observation Codes in the

245 protocol) prior to measurement of the NRU of the cells.

246

247 2.1.3   Measurement of NRU for both 3T3 and NHK Test Methods

248 The treatment medium is removed from the 96-well plates, the cells are rinsed with

249 phosphate buffered saline (PBS), 250 μL NR dye medium is added to the wells (25 μg

250 NR/mL concentration for 3T3 cells, 33 μg NR/mL concentration for NHK cells), and the

251 plates are incubated (37°C ± 1°C, 90% ± 5% humidity, and 5.0% ± 1% CO_2/air) for three

252 hours. After incubation, the NR medium is removed, the cells are rinsed with PBS, and the

253 desorb solution is applied. The plates are shaken on a microtiter plate shaker for 20 to 45

254 minutes to extract NR from the cells and form a homogeneous solution. The absorption (i.e.,

255 OD measurement) of the resulting colored solution is measured (within 60 minutes of adding

256 the desorb solution) at 540 nm ± 10 nm in a spectrophotometric microtiter plate reader, using

257 the blanks as reference. Data from the plate reader is transferred to a Microsoft[®] EXCEL[®]

258 (Microsoft Corporation, Redmond, WA, USA) spreadsheet template (hereafter know as

259 EXCEL[®] template) designed by the SMT and laboratories for statistical analyses for this

260 study.

261

262 **2.2   Descriptions and Rationales of the 3T3 and NHK NRU Test Methods**

263

264 The protocols used in Phases I, II, and III of the validation study (**Appendices B and C**) are

265 modifications of the protocols reported in the *Guidance Document* (ICCVAM 2001b,

266 **Appendix D**). The SMT and the cytotoxicity laboratories provided comments and

267 recommendations in the development of these protocols. The following information is

268 specific to the NICEATM/ECVAM validation study.

269

270 2.2.1   Materials, Equipment, and Supplies

271 *3T3 Cells*

272 3T3 cells (see **Section 2.1.1**), an immortalized mouse fibroblast cell line, were procured from

273 the ATCC by IIVS at passage number 64. IIVS placed the cells in culture to expand the

274   number of cells and cryogenically-preserved them as a pool at passage number 69. ECBC

275   and FAL received frozen ampules of cells at passage number 69 from IIVS, propagated the

276   cells, and cryopreserved multiple ampules of cells at a slightly higher passage number to

277   establish a working cell bank (for each laboratory) for use throughout the study.

278

279   *NHK Cells*

280   These normal human epidermal keratinocytes are primary neonatal foreskin cells pooled

281   from several donors and were obtained from CAMBREX Bio Science Walkersville, Inc. (**see**

282   **Section 2.1.2**). IIVS reserved the specific lot of pooled cells (stored at CAMBREX) for use

283   throughout the study by all laboratories. At each laboratory, cryopreserved NHK cells are

284   thawed from a cryogenic ampule, seeded into culture flasks, propagated according to

285   protocol, then trypsinized and seeded into 96-well plates. NHK cells are passaged only once

286   (to the 96-well plates) and each new assay begins with fresh cells from the cryogenically

287   preserved working bank if NHK cells in the culture flasks are too confluent according to

288   protocol guidelines.

289

290   *Tissue Culture Materials and Supplies*

291   The 3T3 and NHK NRU test methods require general tissue culture materials and supplies

292   (see **Appendices B-1 and B-2** [protocols] for formulations and concentrations of solutions

293   and media). Both test methods use the same materials for solubility testing (**Section 2.8.1**).

294   Freshney (2000) provides information on all aspects of cell culture including materials,

295   supplies, and equipment needed. The following materials are needed for both test methods:

296   • trypsin (i.e., 0.05% trypsin)

297   • PBS

298   • Hanks' Balanced Salt Solution (HBSS) without Ca^{2+} and Mg^{2+}

299   • NR dye

300   • glacial acetic acid

301   • dimethyl sulfoxide (DMSO)

302   • ethanol (ETOH)

303   • distilled water

304

305

306   *Culture Medium*

307   Medium for 3T3 cells consists of DMEM containing high glucose (4.5 gm/L) and

308   supplemented with non heat-inactivated NCS, L-glutamine, penicillin, and streptomycin.

309   The culture medium for NHK cells consists of Clonetics® keratinocyte basal medium

310   (KBM®) supplemented with KBM® SingleQuots® (epidermal growth factor, insulin,

311   hydrocortisone, antimicrobial agents, bovine pituitary extract) and Calcium SingleQuots®

312   (calcium)[all from CAMBREX Corporation].

313

314   *Cell Culture Materials*

315   Laboratory items needed include the following:

316   • sterile, disposable tissue culture plasticware (e.g., 25 cm^2 - 75 cm^2 flasks,

317   multiwell/microtiter plates [96-well], petri dishes) `

318   • cryogenic ampules

319   • pipettes, pipette tips

320   • multichannel solution reservoirs

321   • centrifuge tubes

322   • microporous sterilization filters

323   • general plastic containers

324   • glass tubes (for preparation of reference substance dilutions)

325

326   *Equipment*

327   Performance of the NRU test methods requires a laboratory equipped with a designated cell

328   culture area. Essential equipment for cell culture work and the NRU test method includes:

329   • incubator (37ºC ± 1ºC, 90% ± 5% humidity, 5.0% ± 1% CO_2/air)

330   • laminar flow clean bench/cabinet (standard: "biological hazard")

331   • water bath (37ºC ± 1ºC)

332   • inverted phase contrast microscope

333   • centrifuge (capable of 220 x g)

334   • laboratory balance (capable of measuring to 10 mg)

335   • 96-well plate spectrophotometer (i.e., microtiter plate reader) equipped with 540

336   nm ± 10 nm filter

337   • shaker for microtiter plates

338   • cell counter or hemocytometer

339   • pipetting aid

340   • pipettes, pipettors (multi-channel and single channel, multichannel repeater

341   pipette)

342   • waterbath sonicator

343   • refrigerator

344   • freezer

345   • cryostorage container (liquid nitrogen).

346   • magnetic stirrer

347   • antistatic bar ionizer

348   • personal computer

349   • osmometer

350   • pH meter

351

352   2.2.2   Reference Substance Concentrations/Dose Selection

353   Each laboratory prepares the reference substance immediately prior to testing (i.e., same day

354   as test). Bulk solutions are not prepared for subsequent testing. The highest concentration of

355   dissolved reference substance is identified using the solubility protocol and designated as the

356   2X stock solution. All reference substance dilutions for the assay are serially derived from

357   the stock solution (see **Appendix D** [*Guidance Document*] for serial dilution methods).

358

359   *Range Finder Test*

360   A range finder test is the initial 3T3 and/or NHK NRU test method performed to determine

361   starting doses for the main (definitive) test. The range finder test uses eight concentrations of

362   the reference substance prepared by diluting the stock solution in log dilutions to cover a

363   large concentration range. The highest concentrations applied to the cells are 10 mg/mL for

364   reference substances dissolved in culture medium and 1 mg/mL in medium for reference

365   substances dissolved in DMSO, unless precluded by the solubility of the reference substance.

366   ETOH was not used as a solvent in NRU test methods for any of the 72 reference substances

367   in the NICEATM/ECVAM study.

368

369   If a range finder test does not generate enough cytotoxicity, then a second range finder test is

370   conducted at higher doses, unless precluded by solubility. If solubility is an issue, then more

371   stringent solubility procedures are employed to increase the stock concentration (to the

372   maximum concentration specified in **Appendices B-1 and B-2**). If the test produces a

373   biphasic response curve for NR uptake, then the doses selected for the subsequent definitive

374   tests (concentration-response assays) cover the most toxic dose-response range that includes

375   the range where 50% toxicity is first exceeded (see **Section 2.6.3** – *Unusual Dose-Response*

376   *Curves*).

377

378   *Definitive Test*

379   In the following, because of its capacity to determine the IC_{50} value of a test compound, the

380   main test of the *3T3 and/or NHK* NRU test method will be referred to as the definitive test.

381   The concentration closest to the calculated IC_{50} value in the range finder test can serve as the

382   midpoint of the eight concentrations tested in a definitive test. In the absence of other

383   information (e.g., knowledge of slope for the toxicity curve), the recommended dilution

384   factor is 1.47 ($^{6}\sqrt{10}$), which divides a log into six equidistant steps (e.g., 10, 14.7, 21.5, 31.6,

385   46.4, 68.1, 100), as a starting dilution series. A progression factor of 1.21 ($^{12}\sqrt{10}$) is regarded

386   the smallest factor achievable and was the lowest dosing interval allowed in the validation

387   study. The positive control chemical is tested similarly to the reference substances in the

388   definitive test.

389

390   A successful definitive test is one that meets all of the test acceptance criteria as outlined in

391   the protocol. Definitive tests were repeated as per the protocols if the test failed to meet all

392   test criteria. **Section 2.5** addresses the basis for replicate testing.

393

394   If minimal or no cytotoxicity is measured in the dose range finding test, the maximum dose

395   for a definitive test is as follows:

396   • Reference Substances Prepared in NHK or 3T3 Medium: the highest reference
397   substance concentration applied to the cells in the definitive test is either 100
398   mg/mL (using 200 mg/mL 2X stock) or the maximum soluble dose. A review
399   of the RC chemicals used in this study showed that, among water-soluble
400   chemicals, glycerol had the highest reported IC_{50} value (57 mg/mL). To capture
401   this value during testing and that of other relatively non-toxic chemicals, the
402   100 mg/mL upper concentration limit was established.

403

404   • Reference Substances Prepared in DMSO: the highest test article concentration
405   applied to the cells in the definitive test is either 2.5 mg/mL, or the maximum
406   soluble dose.

407

408   2.2.3   NRU Endpoints Measured
409   *Neutral Red Uptake and Measurement*
410   After cells are exposed to the reference substance or the positive control chemical for the
411   specified period, 3T3 or NHK cells are incubated with the NR dye for three hours, the dye is
412   eluted from the lysosomes using a desorb solution, and the OD of the resulting colorimetric
413   endpoint is measured using a spectrophotometric microtiter plate reader. The OD values are
414   a reflection of the NRU by the cells. The greater the OD value is, the greater the NRU and
415   the higher the percent viability[2] of the cells is in reference to the vehicle control (VC) wells.
416   These OD data are transferred to the EXCEL® template. The mean OD values of the six
417   replicate values (six wells [minimum of four] in the 96-well plate) per test concentration are
418   used to determine relative cell viability by calculating its percentage of the mean NRU of all
419   VC values on the same plate.

420

421   *Determination of $IC_{50,}$ $IC_{20,}$ and IC_{80} Values*
422   The IC_{50} values are determined from the concentration response using a Hill function which
423   is a four parameter logistic mathematical model relating the concentration of the reference
424   substance to the response (typically following a sigmoidal shape). Information on

---

[2] Vehicle control wells are considered to have 100% cell viability (i.e., all cells are alive). Cell viability in other test wells is referenced to the vehicle control value.

425   modifications to the Hill function used in later phases of the validation study may be found in

426   **Section 2.6.3**.

427

428   Data from the EXCEL[®] template were transferred to a template designed by the SMT for a

429   commercially available statistical software program (GraphPad PRISM[®] 3.0, GraphPad

430   Software, Inc., San Diego, CA, USA – hereafter known as PRISM[®] template) to generate the

431   inhibitory concentrations IC_{50}, IC_{20}, and IC_{80} reported as μg/mL of reference substance in

432   solution. IC_{20} and IC_{80} data were collected for potential use in designing a prediction model

433   for estimating human lethal blood concentrations.

434

435   2.2.4   Duration of Reference Substance Exposure

436   The SMT and laboratory representatives reevaluated the reference substance exposure

437   duration recommended in the *Guidance Document* (ICCVAM 2001b) before initiating the

438   NICEATM/ECVAM study. The *Guidance Document* recommends an exposure of 24 hours

439   for the 3T3 cells and 48 hours for the NHK cells. The results from a cytotoxicity study by

440   Riddell et al. (1986) show large differences in cytotoxicity in 3T3 cells induced by some

441   chemicals depending on whether an exposure duration of 24 or 72 hours was used. IIVS

442   conducted studies to evaluate the effect of exposure duration (24, 48, and 72 hours) on the

443   sensitivity of 3T3 cells to six chemicals selected from the list in Riddell 1986. Since the

444   closest fit to the RC regression line (Halle 2003) occurred when 48-hour exposure duration

445   was used, this exposure duration is used in the standardized protocol for 3T3 cells (see

446   **Appendix E**). In addition, IIVS evaluated the sensitivity of NHK cells to the same six

447   chemical using exposure durations of 48 and 72 hours. To make a comparison with the RC

448   regression, the 11 chemicals recommended by the *Guidance Document* were tested in both

449   cell types using the same exposure durations. IIVS scientists concluded that the optimum

450   exposure duration for both cell types was 48 hours (Curren et al. 2003). The SMT concurred

451   and revised the exposure duration in the 3T3 protocol to 48 hours.

452

453

453   2.2.5   <u>Known Limits of Use</u>

454   *Solubility/Volatility*

455   *In vitro* cytotoxicity test methods are inadequate for substances that cannot be dissolved in

456   media, DMSO, or ETOH at a sufficiently high concentration to induce cytotoxicity in excess

457   of 50%. Some reference substance dilutions in this study had precipitates in various 2X

458   concentrations prior to dilution for application to the test plates. Precipitates were observed

459   in a number of test plates after addition of solutions to the cultures and at the end of testing

460   (1X solutions [see **Section 3.5** and **Table 5-11**]). Volatility was detected for a number of

461   reference substances during the range finder tests by observance of cross contamination of

462   wells (i.e., high cytotoxicity in some VC wells). Some volatility was controlled by using

463   plate sealers during the definitive tests (see **Section 2.6.3** – *Testing Volatile Reference*

464   *Substances*). Plate sealers could be used during the range finder tests if the laboratory

465   suspected that the reference substance might be volatile. However, use of plate sealers

466   requires additional laboratory skills and highly volatile reference substances are difficult to

467   test even with the use of plate sealers. Additionally, some test substances (e.g., organic

468   solvents) may react chemically with the plastic plate sealers. Also, chemicals that are

469   unstable or exothermic in water cannot be adequately tested with these test methods.

470

471   *Biokinetic Determinations*

472   The Workshop report (ICCVAM 2001a) provides discussions on the role of the kinetics of a

473   chemical *in vivo* vis a vis its acute systemic toxicity.

474   "Results obtained from *in vitro* studies in general are often not directly applicable to the *in*

475   *vivo* situation. One of the most obvious differences between the situation *in vitro* and *in vivo*

476   is the absence of processes regarding absorption, distribution, metabolism and excretion

477   (i.e., biokinetics) that govern the exposure of the target tissue in the intact organism. The

478   concentrations to which *in vitro* systems are exposed may not correspond to the actual

479   situation at the target tissue after *in vivo* exposure. In addition, the occurrence of metabolic

480   activation and/or saturation of specific metabolic pathways or absorption and elimination

481   mechanisms may also become relevant for the toxicity of a compound *in vivo*. This may

482   lead to misinterpretation of *in vitro* data if such information is not taken into account.

483   Therefore, predictive studies on biological activity of compounds require the integration of

484   data on the mechanisms of action with data on biokinetic behavior."

485

486   Biokinetic determinations were not specifically addressed in this study.

487

488   *Organ-Specific Toxicity*

489   The Workshop report also addresses concerns about which *in vitro* test methods can

490   adequately predict organ-specific toxicity and identifies the organ systems in which failure

491   after acute exposure could lead to lethality (liver, central nervous system, kidney, heart, lung,

492   and hematopoietic system). Each system is reviewed individually and a five-step *in vitro*

493   testing scheme (as opposed to a single *in vitro* test method) that could act as a test battery that

494   may eventually be used as a replacement for *in vivo* acute toxicity testing is proposed.

495   • Step 1 of the proposed *in vitro* scheme recommends performing a physico-

496   chemical characterization and biokinetic modeling.

497   • Step 2 promotes the use of a basal cytotoxicity test method (e.g., 3T3 and NHK

498   NRU test methods).

499   • Step 3 calls for a test to determine the potential that metabolism will mediate the

500   basal cytotoxicity effect.

501   • Step 4 is to assess the test substance's effect on energy metabolism.

502   • Step 5 is to assess the ability of the substance to disrupt epithelial cell barrier

503   function (ICCVAM 2001a).

504

505   Organ-specific toxicity and metabolic effects were not tested in this study.

506

507   2.2.6   Nature of Response Assessed

508   Neutral red is a weakly cationic, water-soluble dye that stains living cells by readily diffusing

509   through the plasma membrane and concentrating in lysosomes. The intensity of the dye in

510   culture is directly proportional to the number of living cells. In addition, since altering the

511   cell surface or the lysosomal membrane by a toxicological agent causes lysosomal fragility

512   and other adverse changes that gradually become irreversible, cell death and/or inhibition of

513   cell growth decreases the amount of neutral red taken up by the culture (see **Section 1.3.1**).

514

515   2.2.7   Appropriate Vehicle, Positive, and Negative Controls

516   *Positive Control (PC)*

517   The *Guidance Document* recommended sodium lauryl sulfate (SLS, Chemical Abstracts

518   Service Reference Number [CASRN] 151-21-3) as an appropriate PC chemical for *in vitro*

519   cytotoxicity test methods (ICCVAM 2001b). SLS is frequently used for this purpose and

520   historical data are available (e.g., Spielmann et al. 1991). A PC test plate was included with

521   each run of any 3T3 and/or NHK NRU test method assay and was treated the same as any

522   reference substance assay plate.

523

524   The acceptable range for the PC IC_{50} was based on the statistical approach recommended in

525   the *Guidance Document*. Initially, in Phase Ia of the validation study, the 3T3 and NHK tests

526   were considered acceptable if the IC_{50} was within the 95% confidence interval of an

527   historical mean IC_{50} value. The SMT decided that the test acceptance criterion for the IC_{50}

528   for Phase III of the validation study (for both cell types) was 2.5 standard deviations of the

529   mean SLS IC_{50} data obtained during Phases I and II. The exception to this was the FAL

530   NHK data, where only the Phase II data were used as the basis for establishing the acceptable

531   range for the PC. SLS data produced at FAL during Phase I was not used due to a protocol

532   change in culturing the cells (see **Section 2.6.2** – *Resultant protocol changes for Phase II*).

533   The historical mean, standard deviation, and acceptance limits were determined separately

534   for each laboratory (see **Table 5-2**).

535

536   *Vehicle Control (VC)*

537   For the NICEATM/ECVAM validation study, the VC consisted of complete DMEM (see

538   **Appendix B-1**) for 3T3 cells and complete Clonetics® KBM® (see **Appendix B-2**) for NHK

539   cells for reference substances dissolved in medium. For reference substances dissolved in

540   DMSO, the VC consisted of medium with the same amount of solvent as that used in the

541   reference substance concentrations that are applied to the 96-well test plate (i.e., 0.5 % [v/v]).

542

543   *Negative Control*

544   A negative control was not incorporated into the NRU test methods. The SMT and study

545   directors decided that the vehicle control would be used in place of a negative control.

546

547   2.2.8   Acceptable Ranges of Control Responses

548   The *Guidance Document* established the use of the <u>absolute value</u> of the OD_{540} value of NRU

549   obtained in the untreated VC to indicate whether the cells seeded in the 96-well plate have

550   grown exponentially with a normal doubling time during the assay. A mean $OD_{540} \geq 0.3$ was

551   recommended as the acceptable range of VC responses and was made a test acceptance

552   criterion for both cell types. Protocols for Phases II and III provide a range of OD values for

553   use as guidance in each phase of the study.

554

555   **Table 2-1   Vehicle Control OD_{540} Ranges**

Phase	OD_{540} Range - 3T3	OD_{540} Range - NHK	Notes
Ia	≥ 0.3 and ≤ 1.1	≥ 0.3 and ≤ 1.1	Test Acceptance Criterion
Ib	≥ 0.30 and ≤ 0.80	≥ 0.60 and ≤ 1.70	Test Acceptance Criterion
II	≥ 0.103 and ≤ 0.813	≥ 0.35 and ≤ 1.50	Target Range (not criterion)
III	≥ 0.103 and ≤ 0.813	≥ 0.205 and ≤ 1.645	Target Range (not criterion)

556

557   In Phase III, 99.5% (914/919) of all 3T3 mean VC OD values and 97% (913/944) of all NHK

558   mean VC OD values were within the target range. Most OD values out of the ranges were

559   from range finding tests and were usually the result of volatile reference substances affecting

560   the VC cells nearest the highest reference substance concentration.

561

562   *VCs as Quality Control*

563   To check for systematic cell seeding errors and potential volatility issues, untreated VCs

564   were placed both at the left side (row 2) and the right side (row 11) of the 96-well plate (see

565   **Appendices B-1 and B-2**). Volatile reference substances generally affect the left side VC

566   (closest to the highest reference substance concentration). The test acceptance criterion was

567   that the left and the right mean of the VCs did not differ by more than 15% from the mean of

568   all VCs. This criterion was used in all phases of the study for reference substances and PC

569   test plates.

570

571   2.2.9   <u>Nature of Experimental Data Collected</u>

572   Each laboratory maintained a Study Workbook to document all aspects of this study. All raw

573   data from cell culture procedures (e.g., cell growth, application of reference substances, NRU

574   test method, etc.) and all solubility studies were recorded in the workbook.

575

576   *NRU OD Measurements*

577   At the conclusion of the NRU desorb step, the OD of the resulting colored solution in each

578   well of the 96-well plates was measured at 540 ± 10 nm in a spectrophotometric microtiter

579   plate reader. Raw OD data from the plate reader was transferred to the EXCEL® template.

580   The template converts the raw data (six wells/reference substance concentration) to derived

581   data by subtracting the mean blank value (two wells/reference substance concentration)

582   associated with each reference substance concentration. The VCs had a total of 12 test wells

583   and 20 blanks. The corrected OD values were referenced to the mean VC OD value and a

584   relative viability (% of VC) was determined for each test well. The percent viability values

585   was then transferred to the PRISM® template for calculation of the IC_{20}, IC_{50}, and IC_{80}

586   values.

587

588   *Type of Data Collected*

589   Originals of the raw data (the Study Workbook and computer printouts of absorbance

590   readings from the plate reader) and copies of other raw data such as instrument logs were

591   collected and archived under the direction of the Study Director according to Good

592   Laboratory Practice (GLP)-compliant procedures.

593

594   The Study Director/technicians entered the following information to the EXCEL® template:

595   •   raw data: OD values from microtiter plate reader

596   •   testing identification for: test facility, chemical code, study number, 96-well

597   plate number, experiment number

598   •   reference substance preparation: solvent used, solvent concentration in dosing

599   solutions, highest stock concentration, dilution factor, pH of 2X dosing

600   solutions, medium clarity/color, presence/absence of precipitate in 2X solutions,

601   PC concentration range

602   •   cell line/type: cell supplier, lot number, cryopreserved passage number, passage

603   number in assay

604   •   cell culture conditions: medium/supplements and supplier and lot numbers,

605   serum concentrations

606   •   test acceptance criteria: acceptable number of values on each side of the IC_{50}

607   (i.e., number of points > 0 and $\leq 50\%$ viability and > 50 and $< 100\%$ viability),

608   acceptable % difference for the VCs, acceptable Hill function R^2 value

609   (coefficient of determination) for the PC, and calculated IC_{50} concentration for

610   the PC

611   •   timeline: dates for cell seeding, dose application, OD_{540} determination

612   •   test results: mean corrected OD_{540} value, Hill function R^2 value, logs of IC_{20},

613   IC_{50}, and IC_{80} (PRISM® template presents data as logs of the IC_x; EXCEL®

614   converts values to IC_x in μg/mL)

615   •   visual observations: protocol codes for cell culture conditions for all reference

616   substance concentrations (i.e., relative level of cell cytotoxicity, cell

617   morphology, presence of precipitate)

618

619   2.2.10   <u>Type of Media for Data Storage</u>

620   Raw data from the NRU cytotoxicity test methods was saved in the EXCEL® template file

621   format provided by the SMT for further analysis of the concentration-response (percent

622   viability calculations). The derived test method data were stored electronically. All

623   EXCEL® and PRISM® files were copied and transferred to compact disks. NICEATM and

624   the laboratories printed copies of all data sheets (stored at NICEATM and at the testing

625   facilities). Copies were also included in the final reports.

626

627   2.2.11   <u>Measures of Variability</u>

628   Each 96-well plate used in the NRU test methods has three main measures of variability.

629   1)   Each plate contains VCs on each end of the plate (columns 2 and 11). The

630   percent difference between each column and the mean of both columns is

631   calculated and was used as a test acceptance criterion. If the difference was

632   greater than 15%, then the test was rejected by the Study Director. This value is

633   an indicator of reference substance volatility and potential cell seeding errors.

634   2) A mean relative viability was determined for each concentration along with the

635   standard deviation and % coefficient of variation (CV).

636   3) Macros were included in the EXCEL® template to perform an outlier test

637   (Dixon and Massey 1981) on data in each well of the test plate. Extreme values

638   at the 99% level were highlighted and could be removed to improve curve fit.

639   The decision as to whether or not to remove outliers was made by the Study

640   Director.

641

642   Other test-to-test measures of variability were considered for this study.

643   •   Each set of assays include a PC plate. If the SLS PC data did not meet test

644   acceptance criteria, then all tests associated with that PC were rejected. The

645   SMT recommended testing a manageable number of definitive test plates (e.g.,

646   4 – 6) with each PC to avoid rejection of reference substance NRU assays that

647   are unacceptable due only to a PC failure. In this validation study, 4.2% of all

648   definitive tests performed were rejected only because the PC failed (i.e., the PC

649   IC_{50} was outside the acceptable confidence limits).

650   •   Standard deviations and CVs were determined for mean IC_{50} values from

651   replicate testing of the same substance. Replicate testing included three

652   definitive tests per reference substance, each performed on a different day.

653

654   2.2.12   Methods for Analyzing NRU Data

655   A calculation of cell viability expressed as NRU was made for each concentration of the

656   reference substance by using the mean NRU of the six replicate values (minimum of four

657   acceptable replicates wells) per test concentration. This value was compared with the mean

658   NRU of all VC values (provided VC values have met the VC acceptance criteria). Relative

659   cell viability was expressed as percent of untreated VC. Raw OD data from the microtiter

660   plate reader was transferred to the EXCEL® template for performance of these calculations.

661   Where possible, the eight concentrations selected for each reference substance tested ranged

662   from no effect up to 100% toxicity.

663

664   The IC_{20}, IC_{50}, and IC_{80} values were determined from the concentration-response by using the

665   PRISM$^{®}$ template and applying a Hill function to the data. The IC_{20} and IC_{80} values were

666   calculated for use in the development of a human prediction model resulting from this study.

667

668   2.2.13   Decision Criteria for Classification of Reference Substances

669   The 3T3 and NHK NRU test methods were not used to classify reference substances in

670   hazard categories but rather to aid in setting the starting dose for acute systemic toxicity

671   assays (i.e., the Up and Down Procedure [UDP], the Acute Toxic Class method [ATC], the

672   Fixed Dose Procedure [FDP]). The RC regression formula (i.e., the prediction model) was

673   used to predict an LD_{50} value from an NRU IC_{50} value. The RC compilation (Halle 2003)

674   contains *in vitro* cytotoxicity information on 347 chemicals (i.e., one average IC_{50x}

675   value/chemical based on multiple reports in the literature) with corresponding *in vivo* acute

676   oral LD_{50} values (mmol/kg) for rats (282 values) or mice (65 values) from RTECS (See Halle

677   2003 for the RC data). **Section 6** addresses the potential of using the *in vitro* NRU

678   cytotoxicity test methods for predicting the GHS hazard category.

679

680   2.2.14   Information and Data Included in the Test Report

681   *Test and Control Substances*

682   (Laboratories in this study worked only with coded reference substances and could

683   not know the specific reference substance information.)

684   • chemical name(s) such as the structural name used by the CASRN, followed by

685   other names, if known

686   • the CASRN, if known

687   • formula weight, if known

688   • purity and composition of the substance or preparation (in percentage(s) by

689   weight)

690   • physicochemical properties (e.g., physical state, volatility, pH, stability,

691   chemical class, water solubility)

692   • treatment of the test/control substances (solubility efforts) prior to testing, if

693   applicable (e.g., vortexing, sonication, warming, grinding)

694   • stability, if known

695   *Information Concerning the Sponsor and the Test Facility*

696   • name and address of the sponsor, test facilities, study director, and laboratory

697   technicians

698   • justification of the test method and protocol used

699   *Test Method Integrity*

700   • the procedure used to ensure the integrity (i.e., accuracy and reliability) of the

701   test method over time (e.g., use of the PC data)

702   *Criteria for an Acceptable Test*

703   • acceptable VC differences (between each column and the mean of both

704   columns)

705   • acceptable concurrent PC ranges based on historical data

706   • number of cytotoxicity points on either side of the IC_{50} (i.e., number of points >

707   0 and ≤ 50% viability and > 50 and < 100% viability)

708   *Test Conditions*

709   • experimental start and completion dates

710   • details of test procedure used

711   • test concentration(s) used

712   • cell type used

713   • description of any modifications of the test procedure

714   • reference to historical data of the model (e.g., solvent and positive controls)

715   • description of evaluation criteria used

716   *Results*

717   • tabulation of data from individual test samples (e.g., IC_{50} values for the

718   reference substance and the PC, reported in tabular form, including data from

719   replicate repeat experiments as appropriate, and means and the standard

720   deviation for each experiment)

721   *Description of Other Effects Observed*

722   • for example, cell morphology, precipitate, NR crystals

723   *Discussion of the Results*

724   *Conclusion*

725 *Quality Assurance (QA) Statement for GLP-Compliant Studies*

726 • This statement indicates all inspections made during the study, and the dates any

727 results were reported to the Study Director. This statement also serves to

728 confirm that the final report reflects the raw data.

729

730 During this study, testing at IIVS and ECBC, the GLP-compliant laboratories, followed

731 additional reporting requirements provided in the relevant guidelines (e.g., OECD 1998; EPA

732 2003a, 2003b; FDA 2003).

733

734 Standard forms for data collection, EXCEL® and PRISM® templates, were developed by the

735 SMT and laboratories. The solubility test form was derived from a standard form provided

736 by IIVS. The EXCEL® template was an adaptation of a template format presented in the

737 *Guidance Document*.

738

739 **2.3   Basis for Selection of the *In Vitro* NRU Cytotoxicity Test Methods**

740

741 As stated in **Section 1**, Workshop 2000 participants recommended that the approach

742 proposed by ZEBET (Halle 1998; Spielmann et al. 1999) be used for rapid adoption so that

743 data could be generated to establish its usefulness with a large number of chemicals

744 (ICCVAM 2001a). To assist in the adoption and implementation of the ZEBET approach,

745 several workshop participants wrote the *Guidance Document* (ICCVAM 2001b). NICEATM

746 and ECVAM used this document as the basis of test method protocol development and

747 designed the validation study to evaluate the performance of the 3T3 and NHK NRU test

748 methods.

749

750 2.3.1   *Guidance Document* Rationale for Selection of *In Vitro* NRU Cytotoxicity Test

751   Methods

752 The *Guidance Document* (ICCVAM 2001b) provides basic protocols for using *in vitro* NRU

753 basal cytotoxicity test methods as the means to predict a starting dose for *in vivo* acute

754 lethality assays. The protocols take advantage of the relationship between *in vitro* IC_{50x}

755 values and *in vivo* LD_{50} values derived from the RC for 347 chemicals (Halle and Spielmann

756   1992; Halle 2003). The 3T3 NRU and NHK NRU test method protocols used in the

757   NICEATM/ECVAM validation study were derived from the document. Guidance was also

758   provided for qualifying these tests for use with the RC regression to predict the starting dose.

759

760   The 3T3 NRU test method has been used most frequently in formal validation programs, all

761   of which were aimed at evaluation of cytotoxicity in predicting eye irritancy. Large-scale

762   studies include Phases I, II, and III of the Cosmetic, Toiletry, and Fragrance Association

763   (CTFA) validation program (Gettings et al. 1991, 1992, 1994a, 1994b); the German eye

764   irritation validation study (Spielmann et al. 1991, 1993, 1996); the European

765   Commission/British Home Office (EC/HO) eye irritation validation study (Balls et al. 1995);

766   and the European Cosmetic Toiletry and Perfumery Association (COLIPA) eye irritation

767   study (Brantom et al. 1997). The 3T3 NRU Phototoxicity Test, a modification of the 3T3

768   NRU test, has been fully validated (Spielmann et al. 1998a,b), and has gained regulatory

769   acceptance. See **Section 9** for comparison of these studies to this validation study.

770

771   2.3.2   *Guidance Document* Rationale for Selection of Cell Types

772   The Workshop (ICCVAM 2001a) concluded that there are no significant differences between

773   the basal cytotoxicity results obtained using permanent mammalian cell lines, primary human

774   cells, or using the IC_{50x} approach of Halle and Spielmann (Halle 2003; Spielmann et al. 1999;

775   Halle and Spielmann 1992). The Workshop recommended that near-term *in vitro* studies

776   designed to reduce and refine animal testing in acute lethality tests should follow the ZEBET

777   approach of using basal cytotoxicity assays in conjunction with the RC database. This can be

778   one of the factors used to identify appropriate starting doses for *in vivo* acute lethality studies,

779   as described by Spielmann et al. (1999).

780

781   *Cell Types for Basal Cytotoxicity Testing*

782   Established rodent (rat and mouse) cell lines were recommended because:

783   •   it was assumed that such cells would give the best prediction of rodent (rat and

784   mouse) acute lethality

785   •   the use of an immortalized standard cell line that is easy to grow and readily

786   available for *in vitro* cytotoxicity testing would hasten the generation of a

787   database that can be used to analyze the usefulness of this approach

788

789   Human cells also offer potential advantages. An analysis of the RC rodent (rat and mouse)

790   acute lethality data relative to cytotoxicity data generated using human cell lines in the MEIC

791   program showed that both human and rodent cells were highly correlative (R^2=0.90)

792   (ICCVAM 2001). A long-term advantage of using human cells is that the human cell

793   cytotoxicity data derived from *in vitro* cytotoxicity testing can be added to human toxicity

794   databases to facilitate the development of test methods that may later better predict acute

795   human lethality.

796

797   *Differentiated Cells for Metabolic Capabilities*

798   The *Guidance Document* explained why highly differentiated cells were not used in the basal

799   cytotoxicity assays. Such cells may not give the best prediction of acute lethality for the

800   large variety of chemicals likely to be tested for acute toxicity (Ekwall et al., 1998). For

801   example, to eliminate the possibility of metabolic activation or inactivation of chemicals,

802   neither hepatocyte nor hepatoma cytotoxicity data were included in the RC database. This

803   does not preclude the use of hepatocytes in future studies, however, either to estimate

804   cytotoxicity or to investigate the effect of metabolism or cell-specific toxicity (Seibert et al.,

805   1996). Hepatocytes are essential to investigations of metabolism-mediated toxicity (Seibert

806   et al., 1996).

807

808   The Workshop participants agreed that the current *in vitro* basal cytotoxicity tests do not take

809   into account metabolism-mediated toxicity. Simple predictive systems (*in vitro* or *in silico*)

810   must be developed for early identification of those substances likely to be metabolized to

811   more toxic or less toxic species than the parent chemical (e.g., Fentem et al., 1993; Seibert et

812   al., 1996; Curren et al., 1998; Ekwall et al., 1999). Participants concluded that the available

813   *in vitro* assays require further development to accurately predict acute lethality (i.e., LD_{50}).

814   See **Section 3.3.4 – Metabolism** for metabolic information on the NICEATM/ECVAM

815   reference substances.

816

817   *Historical Testing*

818   Historical data exists for 3T3 cells including data from controlled and blinded validation

819   studies (Gettings et al. 1991, 1992, 1994a, 1994b; Spielmann et al. 1991, 1993, 1996; Balls et

820   al. 1995; Brantom et al. 1997). Human NHK or fibroblasts have also been used in validation

821   studies for basal cytotoxicity test methods with good results (Willshaw et al. 1994; Sina et al.

822   1995; Gettings et al. 1996; Harbell et al. 1997). See **Sections 5**, **6**, **7**, **8**, and **9** for data

823   generated for the NICEATM/ECVAM validation study.

824

825   **2.4   Proprietary Components of the *In Vitro* NRU Cytotoxicity Test Methods**

826

827   The only proprietary components used in these test methods are the NHK cells and the NHK

828   basal culture medium obtained from CAMBREX Clonetics®. All other components are

829   readily available through various scientific product suppliers. The NHK cells consisted of

830   pooled donor primary neo-natal foreskin keratinocytes from an unidentified source. The use

831   of this specific supplier ensured that the laboratories would have access to the same source of

832   keratinocytes throughout the entire validation study. Keratinocytes from other sources are

833   acceptable if they meet the growth requirements identified in the protocols.

834

835   The contents of the NHK basal culture medium are proprietary, but the formulation is based

836   on a commercially available basal medium (MCDB 153 formulation). This medium was

837   chosen since it was recommended by the laboratories for use with the CAMBREX

838   Clonetics® NHK cells and would be available for the laboratories throughout the study.

839   Other media are acceptable for the NRU test methods if they meet the performance standards

840   prescribed in the media prequalification protocol and achieve parity with the CAMBREX

841   Clonetics® products (see **Appendix B-4** and **Section 2.6.3** – *Inadequate Cell Growth in NHK*

842   *Medium*).

843

844   **2.5   Basis for Number of Replicate and Repeat Experiments for the 3T3 and NHK**

845   **NRU Test Methods**

846

847 The NICEATM/ECVAM study protocols required each laboratory to test the reference

848 substances in at least one range finding test using a log dilution factor and in at least three

849 definitive tests on three different days using a smaller dilution factor than used in the range

850 finding test. Assays were performed over a number of days to assess day-to-day variability.

851

852 Laboratories tested each coded reference substance until three definitive tests met the test

853 acceptance criteria. Additional testing was often dictated by:

854 • chemical issues (low toxicity, volatility, insolubility, and precipitation)

855 • PC failure

856 • technical difficulties such as NR crystal formation

857

858 A stopping rule for insoluble reference substances was incorporated into the protocols to

859 prevent infinite retesting:

860 "If the most rigorous solubility procedures have been performed and the assay

861 cannot achieve adequate toxicity to meet the test acceptance criteria after three

862 definitive tests, then the Study Director may end all testing for that particular

863 chemical."

864

865 **2.6   Basis for Modifications to the 3T3 and NHK NRU Test Method Protocols**

866

867 2.6.1   Phase Ia: Laboratory Evaluation Phase

868 All protocol revisions were implemented during Phase Ia unless otherwise stated.

869

870 *NR Dye Crystals*

871 NR dye crystals formed in the 96-well test plates in both NRU test methods when used at 50

872 μg/mL (OD values measured in the blanks increased from ~ 0.05 to 0.10). Troubleshooting

873 efforts explored incubating the NR medium overnight, centrifuging, filtering, and reducing

874 the concentration of NR dye. The laboratories performed tests using a reduced NR

875 concentration of 33 μg/mL. Since there were no differences in results between tests with 50

876 μg/mL and tests with 33 μg/mL NR, the SMT accepted tests with both concentrations.

877   *Protocol Revision:* The NR dye concentration was reduced to 33 μg/mL for both cell

878   types.

879

880   *3T3 Cell Growth*

881   Cell growth for 3T3 cells was slower than expected in that the cells required more time in

882   culture after seeding cells from the cryogenically-preserved pool into culture vessels to

883   obtain the proper density.

884   *Protocol Revision:* 3T3 cells must be passaged 2-3 times after thawing before reference

885   substance application/toxicity evaluation. The protocol also emphasized attainment of

886   the percent cell confluency required for both cell types prior to reference substance

887   application rather than the amount of time in culture.

888

889   *NHK Cell Growth*

890   The NHK cells also had an additional growth problem that manifested as a ring of

891   dead/dying cells around the center of the wells. Troubleshooting efforts included evaluating

892   various brands of 96-well plates and eliminating the change of medium prior to reference

893   substance treatment. All laboratories participated in evaluating the effect of changing (i.e.,

894   refeeding) or not changing (i.e., no refeeding) the medium by performing a small study with

895   SLS, the PC. Tests were performed 1) after refeeding the cells with fresh medium, and 2) by

896   adding SLS to the medium already on the cells. Control ODs were generally higher in the

897   tests in which the medium was not replenished, but SLS sensitivity was unchanged (see

898   **Table 2-2**). The SMT accepted both tests with refeeding and those without refeeding as long

899   as they met the test acceptance criteria.

900   *Protocol Revision:* Step 2 of the NHK NRU test method was eliminated (change of

901   medium prior to addition of reference substance). The volume of medium with cells

902   placed into the 96-well plates was changed from 250 μL/well to 125 μL/well.

903

904

904   **TABLE 2-2   REFEEDING/NO REFEEDING DATA**

	ECBC[1]		IIVS[2]		FAL[3]	
	Refeed	**No Refeed**	**Refeed**	**No Refeed**	**Refeed**	**No Refeed**
Number of Test Plates	4	4	6	6	2	4
Mean Abs. OD (VC)	0.265	0.621	0.885	1.12	1.41	1.24
Standard Deviation (SD)	0.151	0.322	0.057	0.033	0.127	0.430
SLS IC_{50} (µg/mL)	3.33	3.23	3.41	3.49	6.21	8.14
SLS IC_{50} SD	0.47	0.61	0.58	0.39	0.88	0.40

905   [1]*Edgewood Chemical Biological Center*
906   [2]Institute for In Vitro Sciences
907   [3]FRAME Alternatives Laboratory
908

909   The FAL laboratory could not get satisfactory levels of NHK cell adherence to the 80-cm^2

910   culture flasks when seeded with thawed cells (one ampule) from the cryogenically-preserved

911   pool of cells.

912   *Protocol Revision (FAL only):* Culture flasks were to be coated with fibronectin-collagen

913   to promote adherence.

914

915   *OD Limits*

916   VC control OD limits (OD value must be ≥ 0.3 and ≤ 1.1 as related in the protocols) were

917   frequently unattainable in both test methods. Study Directors reported that the cells were

918   adequately responsive and were neither senescent nor 100% confluent. The SMT withdrew

919   the VC control OD limits as a test acceptance criterion.

920   *Protocol Revision for Phase Ib:* OD data from all laboratories, a review of cell

921   responsiveness (i.e., dose response data), and the ability of each test to pass the other

922   acceptance criteria were analyzed for both cell types and new OD ranges were calculated

923   as guidelines for each cell type.

924

925   *Precipitate Formation*

926   During solubility testing, precipitates were occasionally observed in the 3T3 medium but not

927   in the NHK medium at the same reference substance concentrations. Some liquid reference

928   substances (e.g., 2-propanol) caused precipitation in the 3T3 medium only. The precipitates

929   were attributed to the serum in the 3T3 medium rather than insoluble reference substance.

930   *Protocol Revision:* The reference substance was dissolved in 3T3 medium without NCS.

931   Then, for reference substance exposure, the dissolved 2X reference substance was added

932   to medium containing 10% NCS to reach the final 5% NCS and 1X reference substance

933   concentrations.

934

935   *Dilution Factor*

936   Once a range finder test had been performed, the definitive test assays were to be performed

937   using a $^6\sqrt{10} = 1.47$ dilution scheme centered on the IC_{50}. The laboratories sometimes

938   deviated from the protocols and used dilution factors other than the required one. The SMT

939   accepted data generated using dilution factors other than the recommended 1.47 for definitive

940   tests if all other test acceptance criteria were met. The use of smaller dilution factors

941   generally increased the number of points between 10 - 90% viability and the precision of the

942   IC_{50} calculation was improved.

943   *Protocol Revision:* The $^6\sqrt{10} = 1.47$ dilution scheme was presented as a suggestion and

944   was not a criterion for test acceptance after Phase Ia.

945

946   *Test Acceptance Criteria*

947   *The test acceptance criteria for Phase Ia were:*

948   • the IC_{50} for SLS was within the 95% CI of the historical PC mean established

949   by the Test Facility (*not applicable to Phase Ia*)

950   • mean OD values of the left and right VCs (columns 2 and 11 in the 96-well test

951   plate) did not differ by more than 15% from the mean of all VC OD values

952   • at least two calculated cytotoxicity values, one on either side of the IC_{50},

953   between 10 and 90% viability (*added after commencement of Phase Ia*)

954   • Hill function coefficient of determination $R^2 > 0.9$ <u>or</u> $0.8 < R^2 < 0.9$ and curve

955   fit was evaluated on a case by case basis for acceptability by the SMT (*added*

956   *after commencement of Phase Ia*); (note: this determination would be made by

957   the Study Director in non-validation studies)

958   • OD_{540} of VCs (with blank subtracted) was ≥ 0.3 and ≤ 1.1 (*rescinded after*

959   *commencement of Phase Ia*)

960

961   2.6.2   Phase Ib: Laboratory Evaluation Phase

962   *NR Crystal Formation*

963   FAL and ECBC routinely observed NR crystals forming in the 96-well test plates in 3T3

964   assays at 33 µg/mL NR. All laboratories tested 25 and 33 µg/mL NR concentrations and 2-

965   and 3-hour exposure durations to determine which exposure duration would provide optimal

966   NRU without crystal formation. In addition to determining whether NRU had reached a

967   plateau at these concentrations and durations, the laboratories also tested SLS to determine

968   whether sensitivity to SLS differed under these conditions. Crystals were observed only at

969   33 µg/mL NR when present for 3 hours. **Figure 2-2** shows that the average OD results were

970   very similar for the concentrations and durations tested. **Figure 2-3** shows that the SLS IC_{50}

971   was approximately the same at these concentrations and durations. To minimize changes for

972   the Phase III protocol, the SMT and laboratories agreed to use 25 µg/mL NR for three hours

973   in the subsequent protocol revisions for the 3T3 test method. The NR concentration for the

974   NHK NRU test method remained at 33 µg/mL.

975   *Protocol Revision for Phase II*: The NR concentration for the 3T3 NRU test methods was

976   changed to 25 µg/mL NR for the three-hour incubation. Revised methods for preparation

977   of the NR dye solution included filtration of the solution, maintenance of the solution at

978   37ºC, and application of the NR dye solution to the cells within 15 minutes after

979   removing from 37ºC. Cells should be observed during the NR incubation period of the

980   3T3 and/or NHK NRU test method assays to monitor possible crystal formation.

981

982

982   **Figure 2-2   Optical Density with NR Concentration and Duration**
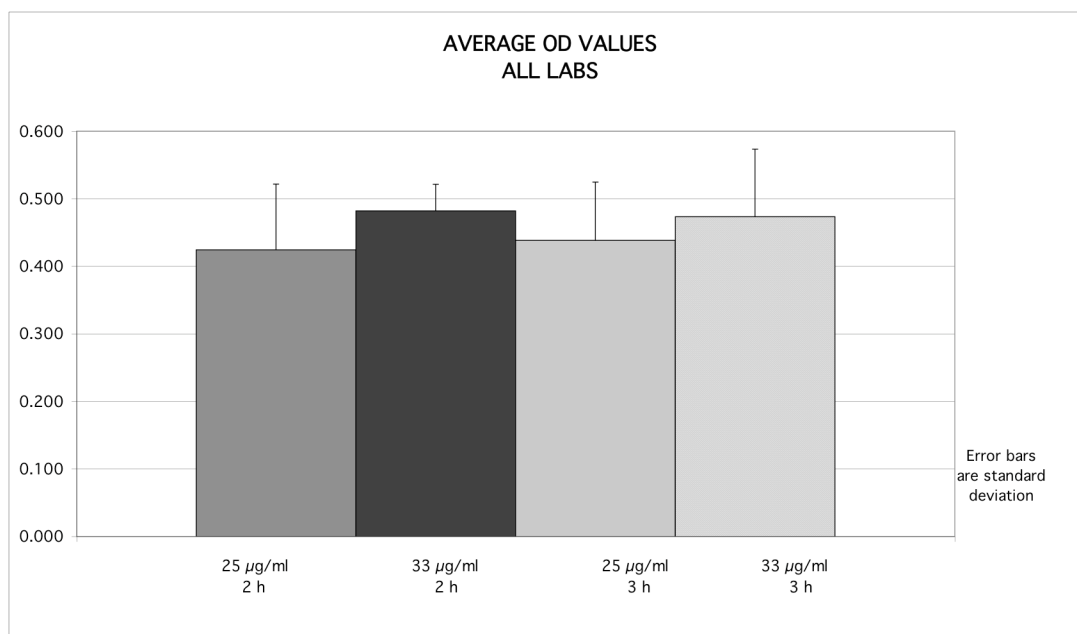
983

984

985   **Figure 2-3   SLS IC$_{50}$ for Each NR Concentration and Duration**
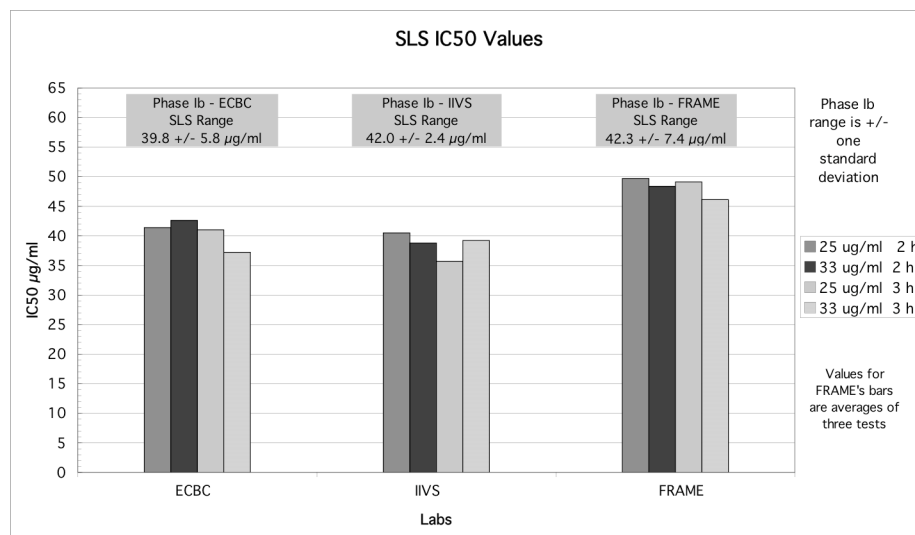
986

987

988   *Heating Reference Substance Solutions*

989   The laboratories had difficulty with the solubility of arsenic trioxide. Mechanical

990   applications for solubilizing reference substances into culture medium were reviewed and

991   revised.

992   *Protocol Revision for Phase II*: The range for duration of heating the reference substance

993   solution was increased from 5 – 10 minutes to 5 – 60 minutes.

994

995   *OD Readings*

996   OD readings were frequently lower than acceptance criteria for the VC wells.

997   *Protocol Revision for Phase II:* The OD range was eliminated as a test acceptance

998   criterion. The OD data from the VCs in the laboratories for both cell types was used to

999   calculate OD ranges to serve as guidelines (see **Section 2.2.9**).

1000

1001   To adjust for potential reference substance interference with NR dye, the reference substance

1002   was added to the blank wells that were used to generate the background OD at 540 nm that

1003   was subtracted from the reference substance concentration ODs. Each reference substance

1004   concentration was applied to six wells containing cells and to two blank wells without cells.

1005

1006   *Laboratory Error Rates*

1007   The SMT suggested that FAL needed additional guidance to become more GLP-like (e.g.,

1008   improve documentation) and to improve performance (i.e., fewer test failures and errors)

1009   throughout Phases Ib and II. The SMT compiled a list of the errors (e.g., transcriptional and

1010   typographical errors in the data sheets) and error rates (number of tests with errors/number of

1011   tests) for the existing Phase Ib data and provided the information to each laboratory (see

1012   **Table 2-3**). IIVS management sponsored a weeklong laboratory training exercise at the IIVS

1013   facilities so that FAL technicians would have exposure to a GLP laboratory environment.

1014   ECBC was invited to participate and all three testing laboratories shared information and

1015   thereby harmonized procedures during the training exercise. Harmonization of the laboratory

1016   procedures illustrated the need to make additional protocol revisions.

1017

1017　**Table 2-3　　Error Rates[a] in Phase Ib by Laboratory and Test Method**

Laboratory	NRU Test Method	
	3T3	**NHK**
ECBC	1/9 (10%)	4/17 (23%)
FAL	42/45 (93%)	12/29 (41%)
IIVS	1/20 (5%)	1/20 (5%)

1018　　　　　　　[a]Number of tests with errors/total number of tests (some data files had more than one error)
1019

1020　*Resultant protocol changes for Phase II*

1021　*The protocol changes include:*

1022　　　• use multi-channel repeater pipettes for plating cells in the 96-well plates,
1023　　　　dispensing plate rinse solutions, NR medium, and desorb solution, but not for
1024　　　　dispensing reference substances to the cells; repeater pipettes are not accurate
1025　　　　enough to deliver equal quantities of the reference substance solution to the
1026　　　　wells

1027　　　• use 8-channel reservoirs for applying dosing solutions to the wells so multi-
1028　　　　channel single delivery pipettes could be used

1029　　　• use a standardized length of time that HBSS rinses remain on the cell
1030　　　　monolayers in flasks during the cell subculturing step

1031　　　• protect plates from high light levels during the shaking step for NR extraction;
1032　　　　all laboratories will cover plates (e.g., with aluminum foil) during this step

1033　　　• allow plates to stand for at least five minutes after the shaking step is complete
1034　　　　and break any bubbles observed in the wells before measuring OD

1035　　　• change the seeding density range for 3T3 NRU test method from 2.5×10^3
1036　　　　cells/well to $2 - 3 \times 10^3$ cells/well

1037　　　• change NHK culture flask size (at FAL) from 80-cm^2 (for start-up of
1038　　　　cryopreserved cells) to 25-cm^2 (same as other laboratories) and discontinue
1039　　　　using a fibronectin-collagen coating

1040

1041　*Test Acceptance Criteria*

1042　*Criteria were modified as follows:*

1043   • the IC$_{50}$ for SLS (PC) is within 2 SDs (approximately 95%) of the historical

1044   mean established by each laboratory in Phase Ia (originally used the 95%

1045   confidence interval)

1046   • mean OD values of the left and right VCs (columns 2 and 11 in the 96-well test

1047   plate) do not differ by more than 15% from the mean of all VC OD values

1048   • at least one calculated cytotoxicity value is between 10 and 50% viability and

1049   one calculated cytotoxicity value between 50 and 90% viability

1050   • Hill function R^2 > 0.9 <u>or</u> 0.8 < R^2 < 0.9 and curve fit is evaluated on a case by

1051   case basis for acceptability by the SMT (note: this determination would be made

1052   by the Study Director in non-validation studies)

1053   • VC OD criteria are based on Phase Ia data (mean ± two SDs): 0.3-0.8 for the

1054   3T3 test method, and 0.6-1.7 for the NHK NRU test method (*rescinded after*

1055   *commencement of Phase Ib*)

1056

1057   2.6.3   <u>Phase II: Laboratory Qualification Phase</u>

1058   All revisions were implemented <u>during</u> Phase II unless otherwise stated.

1059

1060   *Testing Volatile Reference Substances*

1061   When 2-propanol was tested according to the protocol, vapors from the highest concentration

1062   wells contaminated the adjacent VC and appeared to affect some lower concentration wells

1063   (i.e., the wells exhibited unexpectedly reduced levels of NRU). An example dose-response

1064   curve is shown in **Figure 2-4**. The tests for which such contamination was present failed the

1065   VC criterion. When lower concentrations were used to avoid contaminating the VC adjacent

1066   to the highest concentration, toxicity was inadequate to produce an IC$_{50}$. To address this

1067   problem, IIVS repeated their tests using film plate sealers, which isolated all wells from each

1068   other, and obtained acceptable results. Based on these data, the SMT recommended the use

1069   of film plate sealers to the other laboratories to test 2-propanol.

1070

1071   FAL had previous experience using mineral oil as a cell culture cover to keep volatile

1072   reference substances from escaping and provided 2-propanol test data where mineral oil had

1073   been added to each well. The FAL showed that the average oil vs. film IC$_{50}$ values were not

1074   significantly different. However, there was less variability in the film sealer data than the

1075   mineral oil data so the SMT decided on the use of plate sealers.

1076

1077   A general indicator of volatility issues in the NRU test methods is the percent difference in

1078   the mean OD values for the two VC columns on the test plate. If the difference is greater

1079   than 15%, then reference substance volatility is suspected, especially if the VC adjacent to

1080   the highest test concentration had a significantly reduced OD value. Volatility may be an

1081   issue for reference substances with a specific gravity of less than 1. **Table 5-11** lists the

1082   study reference substances that had volatility issues in the NRU test methods.

1083   *Protocol Revision*: The SMT included the use of film sealers to test suspected volatile

1084   compounds in the Phase III protocols.

1085

1086   **Figure 2-4   Representative Dose-Response for 2-propanol in a 3T3 Range Finder**

1087   **Test**

1088
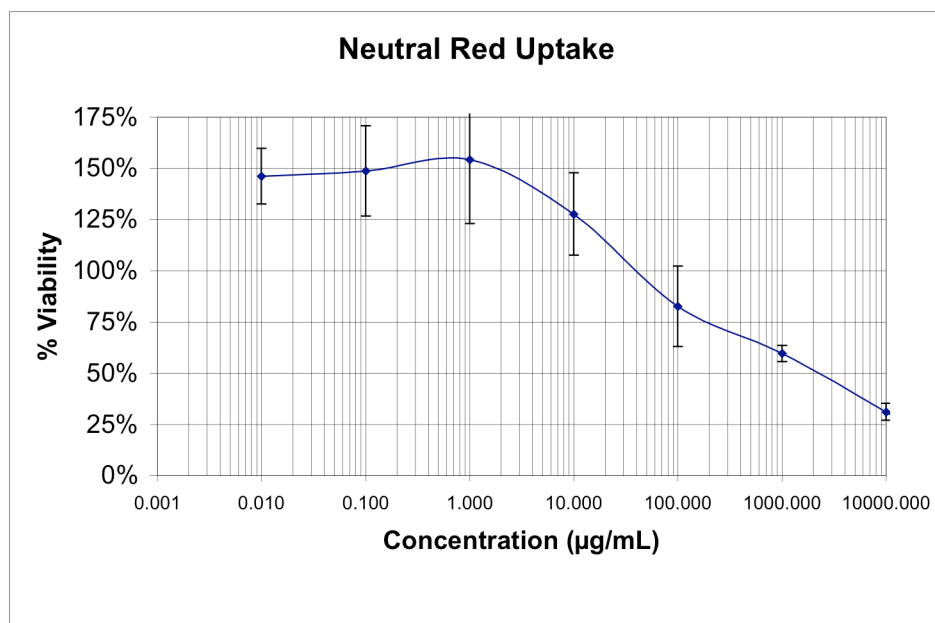
1089   %Difference of the two VC columns from the average VC was 63%. Mean corrected OD
1090   for VC1, adjacent to the highest 2-propanol concentration was 0.070, while that for VC2,
1091   adjacent to the lowest 2-propanol concentration, was 0.310. The 100% viability of the
1092   mean VCs shifted the toxicity curve such that lower concentrations of 2-propanol seem to
1093   have viability percentages much greater than the VCs.
1094

1095

1096   *Unusual Dose-Response Curves*

1097   Some laboratories observed unusual dose-response curves for aminopterin and colchicine.

1098   When the range finder tests produced a biphasic response (see **Figure 2-5** for an example),

1099   the SMT advised the laboratories to focus the definitive tests on the lowest concentrations

1100   that produced responses around 50% viability. In the definitive tests, they noted that no

1101   matter how much reference substance was used, viability was not reduced to 0% (see **Figure**

1102   **2-6**). This effect with colchicine was very reproducible across laboratories in the NHK NRU

1103   test method, but only FAL achieved this type of response with colchicine in the 3T3 NRU

1104   test method. Aminopterin produced a similar dose response in the NHK NRU test method at

1105   ECBC and FAL, but not at IIVS. In the 3T3 NRU test method, only FAL obtained an

1106   unusual response with aminopterin.

1107

1108   The SMT assumed the unusual dose-responses with these reference substances were due to

1109   their mechanisms of action. Colchicine binds to microtubular protein and interferes with

1110   function of mitotic spindles, which arrests cell division (NLM 2003). Aminopterin blocks

1111   the use of folic acid by the cells, which kills cells during the S phase of the cell cycle by

1112   inhibiting metabolism, RNA production, and protein synthesis (NLM 2002). The variability

1113   of results among the laboratories may be due to cells in the culture populations being in

1114   different cell cycle phases when reference substance was applied to the cultures. Application

1115   of reference substance to the cell systems is based on the cells being at a certain monolayer

1116   confluency that assures the cells are in exponential growth phase. A subjective visual

1117   observation of the cell cultures determines time point 0 for the reference substance exposure

1118   period for the NRU test method.

1119

1120

1120   **Figure 2-5   Representative Dose-Response for Aminopterin in a NHK Range**

1121   **Finder Test**

Neutral Red Uptake

1122

1123   Representative dose-response for aminopterin in a NHK range finder test. Laboratories were
1124   instructed to focus definitive tests (concentration-response assays) on the lowest doses that
1125   produced 50% viability.
1126

1127   **Figure 2-6   Representative Dose-Response for Aminopterin in a NHK Definitive**

1128   **Test**

Neutral Red Uptake

1129
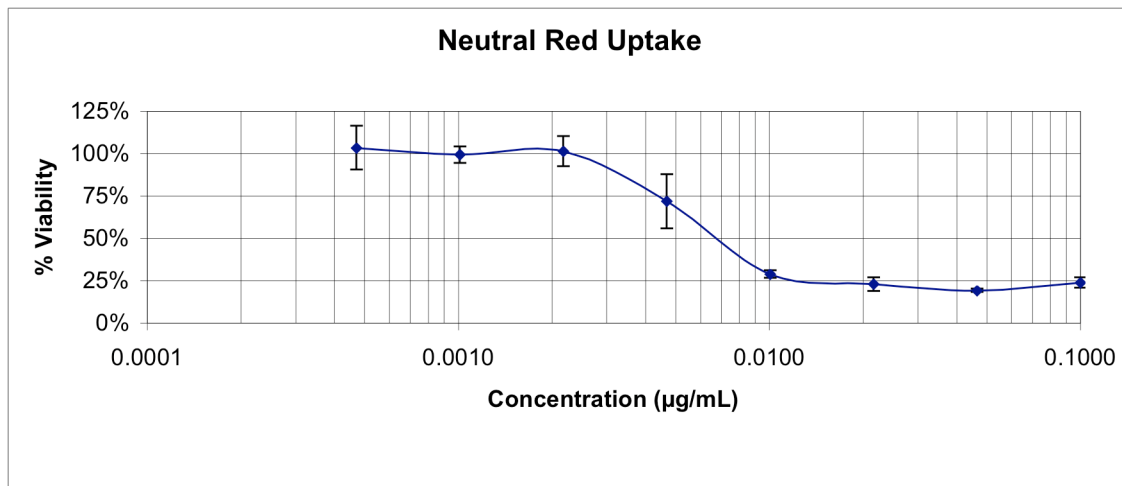
1130   Representative dose-response for aminopterin in a NHK definitive test (concentration-response
1131   assay). %Viability did not reach 0%.

1132

1133

1134

1135

1136   *Hill Function*

1137   The Hill function used in the previous phases of this study was defined as follows:

1138

1139 $$Y = Bottom + \frac{Top - Bottom}{1 + 10^{(logIC50-X)HillSlope}}$$

1140   where Y= response, X is the logarithm of dose (or concentration), Bottom is the minimum

1141   response, Top is the maximum response, $logIC_{50}$ is logarithm of X at the response midway

1142   between Top and Bottom, and HillSlope describes the steepness of the curve.

1143

1144   Since the unusual dose-responses did not fit the Hill function well, R^2 values often failed the

1145   acceptance criterion. To obtain a better model fit, the Bottom parameter was estimated

1146   without constraints (the previous practice was to use Bottom = 0). However, when Bottom $\neq$

1147   0, the EC_{50} reported by the Hill function was not the same as the IC_{50} since the Hill function

1148   relies on EC_{50} defined as the point midway between Top and Bottom. Thus, the Hill function

1149   calculation using the Prism® software was rearranged to calculate the concentration

1150   corresponding to the IC_{50} as follows:

$$X = \log EC_{50} - \frac{\log\left(\frac{Top - Bottom}{Y - Bottom} - 1\right)}{HillSlope}$$

1151

1152   where X is the logarithm of concentration at 50% response, $logEC_{50}$ is logarithm of

1153   concentration at the response midway between Top and Bottom, Top is the maximum

1154   response, Bottom is the minimum response, Y = 50 (i.e., 50% response), and HillSlope

1155   describes the steepness of the curve.

1156

1157   IIVS performed the recalculations for their colchicine tests in the NHK NRU test method, but

1158   the SMT performed the necessary recalculations for the other laboratories. Tests that were

1159   recalculated by the SMT are noted in the data summaries.

1160   *Protocol Revision*: The protocol was revised to state that if a range finding test produces

1161   a biphasic curve, then the concentrations selected for the subsequent tests should cover

1162   the most toxic dose-response range.

1163

1164   *Insoluble Reference Substances*

1165   Lithium carbonate was insoluble in 3T3 medium. Only ECBC was able to expose 3T3 cells

1166   to sufficient lithium carbonate to produce three tests that passed the test acceptance criteria.

1167   Precipitate was reported for two of those tests in the wells at the three highest concentrations.

1168   Since the third highest concentration, 510.2 μg/mL, was approximately the IC_{50} (average was

1169   564 μg/mL), the true IC_{50} for lithium carbonate may actually be lower than that calculated

1170   and therefore the LD_{50} value would be underestimated. The data were not discarded.

1171   *Protocol Revision for Phase III*: The protocol was revised to allow an increase in the

1172   solubility stirring/rocking duration in an incubator from 1 to 3 hours if cytotoxicity in the

1173   range finder test was limited by solubility. Also, a **Stopping Rule for Insoluble**

1174   **Chemicals** was added (see **Section 2.5**)

1175

1176   *Inadequate Cell Growth in NHK Medium*

1177   IIVS and FAL had several NHK NRU test method assay failures that were attributed to poor

1178   cell growth. FAL found that medium/supplement lot combinations that performed poorly in

1179   the NHK NRU test method performed well for the laboratory's research on corneal cell

1180   cultures. The SMT compiled information from the laboratories on the KBM® and

1181   SingleQuot® lot numbers that the laboratories were using along with their assessment of

1182   NHK cell growth. The information was distributed to the laboratories to identify the lots that

1183   produced adequate growth. The SMT also obtained quality assurance and quality control test

1184   results from CAMBREX Clonetics® on the lots of KBM®, but the information provided was

1185   inadequate for determining how the medium would perform in the NHK NRU test method.

1186   *Resolution*: A protocol for prequalifying the medium was developed (see **Appendix B-4**).

1187   For Phase III, the SMT asked IIVS to prequalify new lots of KBM® and SingleQuots®

1188   for use by all laboratories.

1189

1190 *Performance Standards for Media to Support NHK Growth*

1191 A prequalification-of-medium protocol (**Appendix B-4**) was developed and IIVS performed

1192 several tests of different lots of medium and supplements to find various combinations that

1193 maintained the typical growth characteristics of cells in this study. The laboratories then

1194 reserved samples of these acceptable lots at CAMBREX so that the supply of media would

1195 not be interrupted due to unavailability of the materials.

1196

1197 *Test Acceptance Criteria for Prequalifying Media*

1198 - R^2 (coefficient of determination) value calculated for the Hill model fit (i.e.,
1199   from PRISM® software) is ≥ 0.85

1200 - Difference between the mean of all VCs and (a) the left mean VC, and (b) the
1201   right mean VC is $\leq 15\%$

1202 - At least one point > 0 % and ≤ 50.0 % viability and at least one point > 50.0 %
1203   and < 100 % viability

1204 - After meeting all other acceptability criteria, the SLS IC_{50} must be within the
1205   historical range established by the laboratory (i.e., mean SLS $IC_{50} \pm 2.5$ standard
1206   deviations)

1207

1208 *Other Criteria for Prequalifying Media (for consideration by a Study Director)*

1209 - General culture observations: rate of proliferation; percent confluence; number
1210   of mitotic figures per field; colony formation; distribution of cells; absence or
1211   presence of contamination

1212 - Cell morphology observations should include overall appearance (e.g., good,
1213   fair, poor), and presence of abnormal cells

1214 - Mean corrected $OD_{540-550}$ of the VCs

1215 - Cell morphology and confluence of the VCs at the end of the 48-hour treatment.

1216 - Cell doubling time (determined by the laboratory for first time use of the NRU
1217   test method [prior to testing with SLS])

1218

1219 *Test Acceptance Criteria for Phase II*

1220   • IC_{50} for SLS (PC) is within 2.5 SDs of the historical mean established by the

1221   Test Facility (*Phases Ia and Ib*)

1222   •   Mean OD values of the left and right VCs (columns 2 and 11 in the 96-well test

1223   plate) do not differ by more than 15.0 % from the mean of all VC OD values

1224   (*change in decimal point only*)

1225   •   At least one calculated cytotoxicity value ≥ 10.0 % and ≤ 50.0 % viability and at

1226   least one calculated cytotoxicity value ≥ 50.1 % and ≤ 90.0 % viability (*change*

1227   *in decimal point only*)

1228   • $R^2 ≥ 0.90$. Test fails if $R^2 < 0.80$. If the $R^2 ≥ 0.80$ and < 0.90, the SMT

1229   evaluates the model fit (note: this determination is made by the Study Director

1230   in non-validation studies)

1231

1232   2.6.4   <u>Phase III: Laboratory Testing Phase</u>

1233 The changes below were made in the Phase III protocols as a result of the experience in

1234 Phase II.

1235

1236 *Cytotoxicity Values Around the IC_{50}*

1237 Obtaining at least one calculated cytotoxicity value > 0 % and ≤ 50.0 % viability and at least

1238 one calculated cytotoxicity value > 50.0 % and < 100 % viability may be difficult or

1239 unattainable for reference substances with a steep dose response.

1240   *Protocol Revision*: The test acceptance criterion was qualified so that tests with only one

1241   point between 0 and 100 % were acceptable if the smallest practical dilution factor (i.e.,

1242   1.21) was used **<u>and</u>** all other test acceptance criteria were met.

1243

1244 *Data Analysis Revisions*

1245   *Protocol Revision*: If the lowest toxic concentration calculates to be less than 0%, then

1246   the bottom values for IC calculations are set at zero (0) for the Hill function analysis.

1247

1248   *Protocol Revision*: If a biphasic toxicity curve was obtained, the IC_{80} and IC_{50} were

1249   calculated from the initial toxicity part of the curve (the IC_{20} was not determined).

1250

1251   *Protocol Revision*: The requirement for test articles to fit the Hill equation with $R^2 \geq 0.90$

1252   was rescinded. The Hill equation was used to characterize the reference substance

1253   response curve shape rather than establish acceptance criterion. The PC acceptance

1254   criterion was modified to $R^2 \geq 0.85$.

1255

1256   **2.7   Differences in 3T3 and NHK NRU Test Method Protocols and the *Guidance***

1257   ***Document* Standard Protocols**

1258

1259   2.7.1   <u>Optimization of the *Guidance Document* Protocols Prior to Initiation of the Study</u>

1260   As the NICEATM/ECVAM validation study progressed through Phases I and II, the

1261   protocols provided in the *Guidance Document* (ICCVAM 2001b) were optimized to address

1262   problems that were encountered. Changes to the *Guidance Document* protocols are

1263   described below.

1264

1265   •   3T3 cell seeding density for 96-well plates was increased from 1×10^4 cells/well

1266   to $2.0 - 3.0 \times 10^4$ cells/well to achieve adequate cell growth.

1267   •   The calcium concentration in NHK medium was changed from 0.15 mM to 0.10

1268   mM. The test laboratories had expressed concern that cell differentiation would

1269   occur at the higher concentration and requested a lower concentration.

1270   CAMBREX Clonetics®, the supplier of the NHK cells and NHK medium used

1271   in this study, normally grows NHK cells in 0.15 mM calcium without

1272   differentiation issues. The supplier agreed that the cells would grow well at

1273   0.10 mM but should not be cultured at concentrations < 0.10 mM in order to

1274   avoid morphology and growth changes (CAMBREX technical division,

1275   personal communication).

1276   •   NHK cells were subcultured once (rather than the three passages suggested in

1277   the *Guidance Document*). The laboratories expressed concern about

1278   differentiation occurring in the cells if kept in culture too long.

1279   •   The highest final concentrations of DMSO and ETOH in the culture media were

1280   reduced from 1% to 0.5%. IIVS performed experiments with both cell types to

1281   determine the appropriate solvent concentration to avoid toxicity. 3T3 cells

1282   were tested with ETOH at 0.5, 1, and 2% concentrations and DMSO at 0.1, 0.2,

1283   0.3, 0.4, 0.5, 1, and 2% concentrations. The 0.5% concentrations of both

1284   solvents were chosen as optimal since that concentration of ETOH produced no

1285   toxicity. Although 0.5% DMSO produced slight toxicity (i.e., cells were 91%

1286   viable as compared to the control cells – See **Appendix E**), it was chosen by the

1287   SMT and laboratories as an acceptable trade off between slight toxicity and the

1288   ability to reference substances at higher doses and was used throughout the

1289   study (see Curren et al. 2003). However, ETOH was not used as a solvent in the

1290   NICEATM/ECVAM validation study.

1291   •   The pH of reference substance solutions was not adjusted with NaOH or HCl

1292   regardless if solutions became acidic or basic (optimum mammalian cell culture

1293   pH is ~ 7.4 [Freshney, 2000]) since some of the basal cytotoxicity produced by

1294   these reference substances may be due to pH extremes. See **Appendix F** for pH

1295   values of reference substances in culture medium.

1296   •   The CO_2 concentration in the incubator was reduced from 7.5% *(Guidance*

1297   *Document*) to 5.0% since the laboratories were already set up to use 5% CO_2 (a

1298   typical optimum CO_2 concentration for mammalian cell culture).

1299   •   Washing and fixing the cells with a formaldehyde solution prior to NR elution

1300   from the cells was eliminated. FAL's regulatory waste disposal requirements

1301   concerning formaldehyde were an issue and the NR desorb solution (1% glacial

1302   acetic acid, 50% ETOH, 49% H_2O) adequately fixed the cells to the test plate

1303   (INVITTOX 1991). The SMT and laboratories agreed that the use of

1304   formaldehyde was unnecessary.

1305   •   Reference substance exposure time for 3T3 cells was extended from 24 hours

1306   *(Guidance Document)* to 48 hours (see **Section 2.2.6** and **Appendix E**).

1307   •   Cell culture seeding densities for subculture were provided as guidelines and the

1308   laboratories were given liberty to determine adequate cell densities (see **Table**

1309   **2-4**).

1310

1311

1311   **Table 2-4   Cell Culture Seeding Densities**

Protocol	3T3 cells/cm^2 subculture to flasks	3T3 cells/well 96-well Plate	NHK cells/cm^2 subculture to flasks	NHK cells/well 96-well Plate
Guidance Document	1.25×10^4	2.5×10^3	3.5×10^3	$2 - 2.5 \times 10^3$
Phase Ia	$0.42 - 1.68 \times 10^4$	2.5×10^3	$2.5 - 9 \times 10^3$	$2 - 2.5 \times 10^3$
Phase Ib	$0.42 - 1.68 \times 10^4$	2.5×10^3	$2.5 - 9 \times 10^3$	$2 - 2.5 \times 10^3$
Phase II	$0.42 - 1.68 \times 10^4$	$2 - 3 \times 10^3$	$2.5 - 9 \times 10^3$	$2 - 2.5 \times 10^3$
Phase III	$0.42 - 1.68 \times 10^4$	$2 - 3 \times 10^3$	$2.5 - 9 \times 10^3$	$2 - 2.5 \times 10^3$

1312

1313   2.7.2   Optimization of the *Guidance Document* Protocols During the Study

1314   *Changes in Phase Ia*

1315   • To avoid precipitation of serum components, reference substances were
1316   dissolved in treatment medium without NCS for the 3T3 NRU test method
1317   (*Guidance Document* recommended 10% NCS). The final 5% NCS on cells in
1318   the test plate came from the 50:50 dilution of the treatment medium with the
1319   10% NCS in the routine culture medium (see **Section 2.6.1** – *Precipitate*
1320   *Formation*).

1321   • The volume of NHK medium was reduced from 250 µL per well to 125 µL well
1322   for cell seeding. Culture medium was not removed prior to reference substance
1323   application. Cell death occurred during the refeeding step (see **Section 2.6.1** –
1324   *Cell Growth*).

1325   • To avoid NR crystal formation, NR dye concentrations were reduced from 50
1326   µg/mL to 33 µg/mL (3T3) and 25 µg/mL (NHK) (see **Section 2.6.1** – NR Dye
1327   Crystals).

1328   • The PC test acceptance criterion for the IC$_{50}$ was changed for 3T3 and NHK
1329   cells to historical mean ± 2.5 standard deviations instead of within the
1330   recommended 95% confidence interval of historical mean for 3T3 cells and 2
1331   standard deviations for NHK cells.

1332   • The test acceptance criterion for the mean OD$_{540}$ (> 0.3) of the VC was
1333   eliminated. The study protocols provided an OD$_{540}$ range as a guideline (see
1334   **Table 2-1** and **Section 2.2.9.**).

1335

1336   *Changes in Phase Ib*

1337   • NHK cells were deemed ready for reference substance application when they

1338   reached 20+% confluency rather than the range of 30 – 50% confluency.

1339   Laboratory experience in Phase Ia dictated this change.

1340   • A recommendation for obtaining three cytotoxicity points between 10 and 90%

1341   inhibition of NRU for use as a quality check of the dose responses was changed

1342   to become a test acceptance criterion. The dose response curve had to have at

1343   least one calculated cytotoxicity value ≥ 10.0 % and ≤ 50.0 % viability and at

1344   least one calculated cytotoxicity value ≥ 50.1 % and ≤ 90.0 % viability (see

1345   **Section 2.6.2** – *Test Acceptance Criteria*).

1346   • Instructions for using plate sealers were added to the protocols for testing

1347   volatile reference substances (see **Section 2.6.3** – *Testing Volatile Reference*

1348   *Substances*).

1349

1350   **2.8   Overview of the Solubility Protocol**

1351

1352   The SMT, with assistance from the laboratories, developed a solubility protocol to provide

1353   information to the laboratories to optimize the determination of the most appropriate solvent

1354   to use among three solvents: culture medium, DMSO, and ETOH. Each laboratory tested the

1355   solubility of each reference substance using this protocol and provided the data to the SMT

1356   prior to initiating the cytotoxicity testing of each reference substance. The SMT analyzed the

1357   solubility data provided by BioReliance and each testing laboratory, then assigned the

1358   solvents for each test article for this study. This eliminated potential variability in the NRU

1359   test methods that may have been produced if different solvents had been used for testing the

1360   same substance between laboratories.

1361

1362   The solubility protocol is based on an EPA guideline (EPA 1998) that involves testing for

1363   solubility in a particular solvent, beginning at a relatively high concentration and proceeding

1364   to successively lower concentrations by adding more solvent as necessary for dissolution.

1365   Testing stops when, upon visual observation, the procedure produces a clear solution with no

1366   cloudiness or precipitate. The solubility protocol used by the *in vitro* laboratories during

1367   Phase III required testing reference substances in the various solvents at equivalent reference

1368 substance concentrations applied to the cultures. The solubility flow chart in **Figure 2-7**

1369 shows, for example, that 2 mg/mL medium and 200 mg/mL DMSO or ETOH were

1370 equivalent concentrations since they yielded 1 mg/mL in cell culture. When applied to

1371 cultures, medium was diluted by one-half. The 0.5% [v/v] final concentrations were

1372 achieved by diluting DMSO and ETOH by 200. At each concentration, the following mixing

1373 procedures were employed, as necessary, to completely dissolve the reference substance in

1374 this order: vortex (1–2 minutes); sonication (up to 5 minutes); warming to 37°C (5 – 60

1375 minutes [NRU protocols allow warming to be extended to three hours if cytotoxicity in the

1376 range finder test was limited by solubility]). If the reference substance was still undissolved,

1377 the next concentration/solvent was tested.

1378

1379 **Figure 2-7   Flow Chart for Determination of Reference Substance Solubility in**

1380 **Medium, DMSO, or ETOH**

Tier	1	2	3	4	5
Concentration in 3T3 and NHK Media	**Start Here** 20 mg/mL	2 mg/mL	0.20 mg/mL		
Concentration in DMSO		200 mg/mL	20 mg/mL	2 mg/mL	0.2 mg/mL
Concentration in Ethanol		200 mg/mL	20 mg/mL	2 mg/mL	0.2 mg/mL **End**
Concentration on Cells	10 mg/mL	1 mg/mL	0.1 mg/mL	0.01 mg/mL	0.001 mg/mL

Incomplete solubility arrows connect each tier.

1381

1382 Notes: 3T3 Medium - DMEM (Dulbecco's Modification of Eagle's Medium) with supplements; NHK
1383 medium - KBM® (Keratinocyte Basal Medium) with supplements (from CAMBREX Clonetics®).
1384

1385

1385   **2.9   Components of the Solubility Protocol**

1386

1387   2.9.1   <u>Medium, Supplies, and Equipment Required</u>

1388   *Medium and Chemical Supplies*

1389   •   <u>3T3 Cell Medium</u>: DMEM without L-Glutamine and containing Hanks' salts

1390   and high glucose [4.5gm/l]; L-Glutamine 200 mM; NCS

1391   •   <u>NHK Cell Medium</u>: Keratinocyte Basal Medium without Ca^{++} (KBM[®], Clonetics[®]

1392   CC-3104); KBM[®] SingleQuots[®] medium supplements (Clonetics[®] CC-4131) --

1393   epidermal growth factor, insulin, hydrocortisone, antimicrobial agents, bovine

1394   pituitary extract; Calcium SingleQuots[®] (Clonetics[®] CC-4202);

1395   penicillin/streptomycin solution

1396   •   U.S.P. analytical grade DMSO

1397   •   U.S.P. analytical grade (100%, non-denatured) ETOH

1398

1399   *Equipment*

1400   •   waterbath (37°C)

1401   •   sonication unit

1402   •   vortex unit

1403   •   pippettors (micropipettors)

1404   •   balance

1405   •   pH meter

1406

1407   *Procedures*

1408   The first Phase III solubility protocol procedure was the dissolving of ~10 mg of a reference

1409   substance in ~ 0.5 mL medium (both 3T3 and NHK media were tested) for a concentration of

1410   20 mg/mL (see **Appendices B-1 and B-2**). In order, the mixture was vortexed for 1-2

1411   minutes, sonicated for up to 5 minutes, and warmed to 37°C for 5-60 minutes as necessary to

1412   dissolve the reference substance. The endpoint for dissolution was that a clear, not cloudy

1413   solution with no noticeable precipitate. If the reference substance was not soluble in medium

1414   at 20 mg/mL, then more medium was added to a concentration of 2 mg/mL (i.e., a total

1415   volume of ~5 mL) (Step 2). The mixing procedures were repeated as necessary to dissolve

1416   the reference substance. If the reference substance was not dissolved, ~10 mg reference

1417   substance in ~0.5 mL DMSO was added in an attempt to dissolve it at 200 mg/mL DMSO

1418   (Step 3). If the reference substance was not dissolved, the same concentration was attempted

1419   in ETOH (Step 4). Step 5 began in the same way with 0.2 mg/mL medium and then to 20

1420   mg/mL DMSO and then 20 mg/mL ETOH.

1421

1422   Determination of solubility of reference substances was limited to visual observation of the

1423   reference substance in solution. If a solution appeared clear, then solubility testing ceased. If

1424   particles were visible or the solution appeared cloudy, then more stringent mixing procedures

1425   were employed. If necessary, the solubility procedure proceeded to the next

1426   solvent/concentration tier. The duration of the solubility test was dependent on mechanical

1427   procedures used to achieve solubility. Some reference substances were immediately

1428   solubilized (e.g., liquids) and others required up to 60 minutes of heating and other

1429   mechanical procedures.

1430

1431   2.9.2   <u>Data Collection</u>

1432   All laboratories (including the reference substance distribution laboratory [BioReliance])

1433   used a worksheet designed to capture the solubility information for the reference substances.

1434   The protocol's tiered approach to determining solubility of each reference substance was

1435   followed. The endpoint for each step was a visual observation of the solution and a

1436   documented comment of soluble or insoluble. Each worksheet contained:

1437   • reference substance code and physical description

1438   • solvent (3T3 medium, NHK medium, DMSO, ETOH)

1439   • amount of reference substance (mg)

1440   • volume of solvent added and total volume (mL)

1441   • concentration (µg/mL)

1442   • pH and solvent color

1443   • mechanical procedures (vortexing, sonication, heating)

1444   • comments (soluble/insoluble at the particular concentration; visual

1445   observations)

1446

1447   The solubility test data from the laboratories were transferred via email to the SMT and

1448   stored on the NICEATM server and as hard-copy printouts. Each laboratory also maintained

1449   electronic and hard-copy files of the data.

1450

1451   2.9.3   Variability in Solubility Measurement

1452   Solubility analyses were not replicated since within-laboratory results were not expected to

1453   vary. Comparison of the laboratory results to determine laboratory concordance for the 72

1454   reference substances (see **Section 4** for results) provided a measure of variability among the

1455   laboratories (see **Section 7**).

1456

1457   2.9.4   Solubility and the 3T3 and NHK NRU Test Methods

1458   Reference substance solutions were monitored throughout all aspects of the *in vitro* NRU

1459   cytotoxicity test methods and observations were documented. The 2X and 1X solutions for

1460   the range finder tests were permitted to contain precipitates. The lowest concentration of

1461   reference substance in a 2X solution that contained observable precipitates, particles,

1462   globules, or oily droplets was noted in the EXCEL® template. After reference substance

1463   exposure, all wells of the 96-well test plates were observed microscopically and scored using

1464   a visual observation code as per the NRU protocol. The code addressed growth

1465   characteristics and the presence or absence of precipitates. The Study Directors made

1466   determinations of test acceptance based on the effect that precipitates had on the NRU

1467   results.

1468

1469   2.9.5   Methods for Analyzing Solubility Data

1470   During Phase III, the SMT used the solubility data from all the laboratories to determine the

1471   solvent that would be used for cytotoxicity testing (see **Section 5** for solubility results and

1472   SMT selections). If the solubility of an individual reference substance in 3T3 medium and

1473   NHK medium was different, the SMT chose the same solvent for both test methods, rather

1474   than choosing one for the 3T3 NRU test method and one for the NHK NRU test method. For

1475   example, if solubility in one medium was $\geq$ 2 mg/mL and solubility in the other medium was

1476   < 2 mg/mL, and the reference substance was soluble in DMSO at 200 mg/mL, then the SMT

1477   selected DMSO as the solvent for cytotoxicity testing. Where possible, the SMT chose a

1478   solvent such that all cytotoxicity laboratories could obtain solubility at some concentration.

1479   For example, if a reference substance had low solubility in medium (i.e., 2 mg/mL) at one

1480   laboratory and high solubility in DMSO at the other laboratories, the SMT chose DMSO.

1481

1482   Solubilizing enough reference substance to produce cytotoxicity was challenging for

1483   relatively insoluble low toxicity reference substances such as lithium carbonate (in the 3T3

1484   NRU test method) but generally was not a problem for toxic reference substances. Some

1485   insoluble and highly toxic reference substances were problematic, however, because the

1486   amount of powdered reference substance added to solvent was very small, so it was difficult

1487   to determine the absence of solute particles in solution (i.e., if the solution was visibly clear).

1488   Any undissolved reference substance remaining might have been too little to see. Arsenic

1489   trioxide is an example of such a solute.

1490

1491   **2.10   Basis of the Solubility Protocol**

1492

1493   The solubility protocol used by BioReliance, which tested solubility of the reference

1494   substances prior to testing by the *in vitro* laboratories, is provided in **Appendix G**. The

1495   protocol is based largely on information from the literature and Internet searches for

1496   solubility procedures, the experience of the SMT and IIVS, and the solubility and IC_{50}

1497   information for the RC chemicals (Halle 2003). The only formal solubility protocol

1498   discovered was the EPA Product Properties Test Guideline, OPPTS 830.7840 Water

1499   Solubility Column Elution Method; Shake Flask Method (EPA 1998).

1500

1501   2.10.1   Initial Solubility Protocol Development

1502   BioReliance tested reference substances in cell culture media at 2000 mg/mL, 400 mg/mL,

1503   and 200 mg/mL, and if not soluble, in DMSO, and then ETOH at the same concentrations

1504   (initial protocol). It was apparent that these concentrations were not low enough when the

1505   laboratory was unable to achieve solubility for arsenic trioxide. The solubility protocol was

1506   revised twice to lower the concentrations tested (see **Table 2-5**). An extra tier of

1507   concentrations ≤ 1 mg/mL was added for insoluble reference substances. Because of this

1508   experience, this solubility protocol for the cytotoxicity laboratories was revised to reduce the

1509   number of steps required (by testing in log units) and to test in tiers in which the reference

1510   substance concentrations reflected the same concentrations in cell cultures.

1511

1512   In Phases Ib and II, the SMT used the data from BioReliance to determine the solvent for the

1513   *in vitro* laboratories to use for NRU testing. When it became apparent that the laboratories

1514   sometimes obtained different results than those reported by BioReliance, the SMT used the

1515   cytotoxicity results from all the laboratories to determine the solvents for Phase III reference

1516   substances.

1517

1518   **Table 2-5   Comparison of Concentrations Tested in Various Solubility Protocols**

Solubility Protocol Version	Concentrations Tested (mg/mL)					
	Step 1	Step 2	Step 3	Step 4	Step 5	Steps 6-10
BioReliance 1 (4/26/02) and Phase Ia for cytotoxicity laboratories	2,000	400	200			
BioReliance 2 (9/17/02)	200	40	20	10	2	
BioReliance 3 (10/11/02)	200	40	20	10	2	1, 0.5, 0.25, 0.125, 0.05
Phases Ib, II, III for cytotoxicity laboratories	20 Medium	2 Medium 200 DMSO 200 ETOH	0.2 Medium 20 DMSO 20 ETOH	2 DMSO 2 ETOH	0.2 DMSO 0.2 ETOH	

1519   DMSO – dimethyl sulfoxide
1520   ETOH – ethanol
1521   Medium – cell culture medium

1522

1523   The protocol provided a tiered approach for determining the 2X stock concentration for each

1524   reference substance, based on the solvent and solubility of the reference substance (see

1525   **Figure 2-7**). The solubility protocol was developed to reduce the number of steps for testing

1526   (compared to that used by BioReliance) so that solubility testing was less time consuming

1527   (see **Appendix B-3**).

1528

1529   2.10.2   Basis for Modification of the Phase II Protocol

1530   All three cytotoxicity laboratories found arsenic trioxide (tested in Phase Ib) less soluble than

1531   that reported by BioReliance (0.25 mg/mL in 3T3 medium and 0.05 mg/mL in NHK

1532 medium). Use of the solubility procedures in the protocol did not dissolve arsenic trioxide.

1533 IIVS warmed the stock solution (at least 200 μg/mL for 2X) for longer than the protocol

1534 specified (i.e., 30-50 min) but still had small, undissolved particles persist in the non-

1535 homogeneous stocks (i.e., particles readily fell out of suspension). ECBC obtained a clear

1536 solution (highest 2X concentration was 30-50 μg/mL), but found precipitated particles after

1537 the solution stood at room temperature. Sonication time was increased to 15-30 min, and

1538 heating time to ~ 30 min to get a finer suspension. This procedure achieved a more

1539 homogeneous mixture, resulting in better series dilutions and more uniform application of

1540 reference substance to the cells. FAL stirred the suspension (~ 20-90 μg/mL) in the CO_2

1541 incubator for 1.5 to 2 hours to get clear medium.

1542   *Protocol Revision for Phase II*: The duration of the solution heating range was increased

1543   from 5-20 minutes to 5-60 minutes.

1544

1545 **2.11   Summary**

1546

1547   • The *Guidance Document* NRU protocols were the basis of the

1548   NICEATM/ECVAM study protocols. The SMT and laboratories made initial

1549   modifications to the protocols prior to implementation of the study. Other

1550   protocol modifications were made after commencement of testing and were the

1551   result of comments and recommendations from the laboratories and the SMT.

1552   The resulting optimized protocols were used in the main testing phase (Phase

1553   III) and were the final protocols for the NICEATM/ECVAM study.

1554   • The solubility protocol was developed to provide specific guidance to

1555   laboratories to assure that solubility issues could be satisfactorily addressed and

1556   reference substances from a specific study set could be adequately prepared and

1557   evaluated for *in vitro* cytotoxicity effects.

1558

1558

1559

1560

1561

1562

1563

1564

1565

1566

1567

1568

1569

1570

1571

1572

1573   ***[This Page Intentionally Left Blank]***

1574

27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42   ***[This Page Intentionally Left Blank]***
43
44

44   **3.0   REFERENCE SUBSTANCES USED FOR VALIDATION OF THE 3T3 AND**

45   **NHK NRU TEST METHODS**

46

47   This section discusses the rationale for the selection of the 72 reference substances tested to

48   validate the 3T3 and NHK NRU test methods for determining starting doses for rodent acute

49   oral systemic toxicity testing. Information regarding chemical class and physical/chemical

50   characteristics is provided, as is the available information on toxicological characteristics,

51   such as target, organ, extent of metabolism, and mechanism of action, for the 72 reference

52   substances. Such information may be useful for characterizing the performance of the 3T3

53   and NHK NRU test methods for various chemical types. Chemical supplier and purity

54   information are provided as well as the methods for purchasing, coding, and distributing the

55   substances to the testing laboratories.

56

57   **3.1   Rationale for the Reference Substances Selected for Testing**

58

59   This section describes the procedure used to select the 72 reference substances tested in the

60   NICEATM/ECVAM validation study.

61

62   3.1.1   Reference Substance Selection Criteria

63   The SMT selected reference substances for testing in 2001 and 2002 using a process based

64   primarily on general recommendations made by Workshop 2000 participants (ICCVAM

65   2001b). The following criteria were used:

66   •   the toxicities of the reference substances should be evenly distributed across the

67   expected range of LD_{50} values (i.e., the GHS classification for acute oral

68   toxicity [UN 2005])

69   •   the reference substances should cover a wide range of structural and use classes,

70   according to the needs of various user communities

71   •   substances with human toxicity data and/or human exposure potential (i.e.,

72   substances of interest to society) should be included

73

74   **Table 3-1** shows the GHS classification scheme which classifies chemicals into five toxicity

75   categories or an unclassified group based on acute oral LD_{50} values (UN 2005).

76

77   **Table 3-1   UN GHS[1] Classification Scheme for Acute Oral Toxicity**

Category	LD_{50} (mg/kg)
1	$LD_{50} \leq 5$
2	$5 < LD_{50} \leq 50$
3	$50 < LD_{50} \leq 300$
4	$300 < LD_{50} \leq 2000$
5	$2000 < LD_{50} \leq 5000$
Unclassified	$LD_{50} > 5000$

78   [1]Globally Harmonized System of Classification and Labelling of
79   Chemicals (UN 2005)
80

81   For the purposes of toxicity classification, the rodent oral LD_{50} values for individual

82   reference substances were obtained from readily available toxicological databases. Rat LD_{50}

83   values were preferred, but mouse LD_{50} values were used (three reference substances) when

84   rat data were unavailable. However, mouse data were not used in the regression analyses

85   (See **Section 6**). The Registry of Cytotoxicity (RC) is a database of acute oral LD_{50} values

86   for rats and mice obtained from RTECS® and IC_{50} values from *in vitro* cytotoxicity assays

87   using multiple cell lines and cytotoxicity endpoints for chemicals with known molecular

88   weights (Halle 1998). The toxicological databases, in order of preference, were:

89   •   the RC, which contains LD_{50} values that came largely from the 1983/84

90   RTECS® (Halle et al. 1998)

91   •   the current RTECS® (MDL Information Systems 2001, 2002)

92   •   the current Hazardous Substances Data Bank (HSDB; U.S. National Library of

93   Medicine [NLM] 2001, 2002).

94

95   To assure that a wide range of structural and use classes were selected, reference substances

96   of interest to the various U.S. regulatory agencies, as determined from chemical lists

97   (personal communication) from the agencies, were included. Chemicals with human toxicity

98   data and/or human exposure potential (i.e., chemicals of interest to society) were chosen by

99   mining publicly available databases (e.g., NTP test database) for potential candidates.

100

101   3.1.2   <u>Candidate Reference Substances</u>

102   *Sources of Candidate Chemicals*

103   The process of identifying the 72 reference substances started with the compilation of a

104   database that ultimately contained 116 candidate chemicals. The intent of the SMT was to

105   compile a database with more than 12 chemicals in each toxicity category that also met the

106   other criteria, and then to prioritize the chemicals in each category to select the 72 reference

107   substances to be tested. As recommended by the Workshop 2000 participants (ICCVAM

108   2001a), the following publicly available databases and other indicated sources were used to

109   identify candidate chemicals:

110   • the MEIC program, which collected human toxicity data and *in vitro* toxicity
111   data from 61 test methods for the first 50 chemicals (Ekwall et al. 1998)

112   • the RC (Halle 1998), which contains a compilation of *in vitro* cytotoxicity and
113   *in vivo* rodent LD_{50} data for 347 chemicals

114   • the Toxic Exposure Surveillance System (TESS) (Litovitz et al. 2000), which
115   compiles reports of toxic human exposures from poison control centers
116   throughout the United States

117   • pesticides recommended for consideration by the EPA Office of Pesticide
118   Programs (OPP)

119   • the *Guidance Document* (ICCVAM 2001b), which reported the NRU results for
120   11 RC chemicals using protocols similar to those used in the
121   NICEATM/ECVAM validation study

122   • the U.S. NTP test database, which contains information on the toxicity of
123   chemicals relevant to human exposure (NTP 2002)

124   • the EPA High Production Volume (HPV) Challenge Program, which is a
125   voluntary testing program to provide the public with a complete set of baseline
126   health and environmental effects data for each chemical that is manufactured
127   within or imported into the United States at amounts > 1 million pounds/year
128   (EPA 2000)

129

130   *Selection of Candidate Chemicals*

131   The complete list of candidate chemicals is provided in **Table 3-2.** The left side of **Table 3-2**

132   presents selected chemicals and the right side presents the alternate chemicals. The candidate

133   chemicals are grouped by GHS acute oral toxicity classification. For each candidate

134   chemical, the table provides the corresponding rat or mouse oral LD_{50} value, the database(s)

135   or other source(s) used to identify the chemical as a potential candidate, and the type of

136   product and/or use for the chemical. Product/use categories were identified from HSDB

137   (NLM 2001, 2002) or RTECS® (MDL Information Systems 2001, 2002).

138

139   The final list of candidate chemicals compiled by the SMT included:

140   •   65 MEIC chemicals. These include the first 50 chemicals evaluated by MEIC

141   as well as another 15 chemicals that were identified for future evaluation (C.

142   Clemedson, personal communication 2001). Twenty of these chemicals were

143   identified for the EDIT program, a follow-on project to the MEIC study to

144   develop supplementary toxicity and kinetic tests (to determine distribution of

145   chemicals in the body and biotransformation of chemicals to more toxic

146   metabolites) to improve the prediction of human toxicity by the battery of tests

147   identified as the best predictors in the MEIC program (Clemedson et al. 2002).

148   The EDIT chemicals were selected by excluding MEIC chemicals that were

149   volatile, those that precipitated at the IC_{50} dose level, and those with sparse or

150   insufficient data on human toxicity or mechanism of acute toxicity

151   •   16 pesticides with extensive human exposure nominated by the EPA OPP.

152   These included fenpropathrin, endosulfan, bromoxynil (phenol), fipronil,

153   carbaryl, rotenone, metaldehyde, molinate, 1,3-dichloropropene, dichlorvos,

154   chlorpyrifos, sodium arsenite, triphenyltin hydroxide, cycloheximide, acrolein,

155   and boric acid. Pentachlorophenol was also nominated, but was already on the

156   candidate list since it was a MEIC chemical

157

158

159

Table 3-2   Candidate Chemicals for the 3T3 and NHK NRU Test Methods Validation Study

Selected Chemicals				Alternate Chemicals				
GHS Category[1]/Chemical	Rodent Oral LD$_{50}$[2] (mg/kg)	Source[3]	Product/Use[4]	GHS Category[1]/Chemical	Rodent Oral LD$_{50}$[2] (mg/kg)	Source[3]	Notes[5]	Product/Use[4]
LD$_{50}$ ≤ 5 mg/kg								
Mercury II chloride	1	MEIC, EDIT, RC (outlier), TESS, NTP	Preservative/ manufacturing/ insecticide	Aflatoxin B1	5.0	RC (outlier)	Prohibitively expensive	Food contaminant
Triethylenemelamine	1	RC (outlier), NTP	Manufacturing/ insect sterilant					
Sodium selenate	2[a]	TESS, NTP	Feed additive					
Busulfan	2	RC (outlier), NTP	Pharmaceutical (antineoplastic)					
Cycloheximide	2	RC (outlier), NTP	Antibiotic/fungicide					
Disulfoton	2	RC (outlier), EPA, NTP	Pesticide (insecticide)					
Parathion	2	RC (outlier), EPA, NTP	Pesticide (insecticide)					
Strychnine	2[a]	MEIC, TESS, EPA	Pesticide (rodenticide)					
Aminopterin	3[b]	RC	Pharmaceutical (antineoplastic); Rodenticide					
Phenylthiourea	3	RC (outlier), NTP	Pesticide (rodenticide)					
Epinephrine bitartrate	4[b]	RC (outlier), NTP (HCl salt)	Pharmaceutical (adrenergic)					
Physostigmine	5[a]	EHS	Pharmaceutical (anticholinesterase)					
5 < LD$_{50}$ ≤ 50 mg/kg								
Colchicine	6[b]	MEIC, RC, TESS, NTP	Pharmaceutical (gout suppressant)	2,4-Dinitrophenol	30	RC (outlier), NTP, HPV		Pesticide (fungicide/ insecticide) manufacturing
Potassium cyanide	10	MEIC, EDIT, RC (outlier), TESS	Electroplating	t-Butylamine	44[a]	EPA, NTP, HPV		Manufacturing

Table 3-2   Candidate Chemicals for the 3T3 and NHK NRU Test Methods Validation Study

Selected Chemicals				Alternate Chemicals				
GHS Category[1]/Chemical	Rodent Oral LD$_{50}$[2] (mg/kg)	Source[3]	Product/Use[4]	GHS Category[1]/Chemical	Rodent Oral LD$_{50}$[2] (mg/kg)	Source[3]	Notes[5]	Product/Use[4]
Dichlorvos	17[a]	TESS, EPA, NTP, HPV	Pesticide (insecticide)	Acrolein	46	RC, TESS, EPA, NTP, HPV	Volatile (BP=52°C)	Pesticide (herbicide/ rodenticide/ algicide), manufacturing
Digoxin	18[b]	MEIC, EDIT, RC (outlier), TESS	Pharmaceutical (antiarrhythmic)					
Fenpropathrin	18[a]	EPA	Pesticide (insecticide)					
Endosulfan	18[a]	TESS, EPA, NTP	Pesticide (insecticide)					
Arsenic III trioxide	20	MEIC, EDIT, RC, TESS, EPA, NTP	Pesticide (insecticide)					
Thallium I sulfate	29[b]	MEIC, EDIT, RC (outlier), TESS	Pesticide (rodenticide/ insecticide)					
Sodium arsenite	41[a]	TESS, NTP	Pesticide (herbicide, insecticide, fungicide)					
Triphenyltin hydroxide	44	RC, EPA, NTP, HPV	Pesticide (fungicide/ insecticide)					
Sodium dichromate dihydrate	50	RC, EPA, GD, NTP	Oxidizing agent					
Nicotine	50	MEIC, EDIT, RC (outlier), TESS, EPA, NTP	Pharmaceutical (stimulant)					
50 < LD$_{50}$ ≤ 300 mg/kg								
Paraquat	58	MEIC, EDIT, RC (outlier), TESS, EPA	Pesticide (herbicide)	Pentachlorophenol	51	MEIC, RC (outlier), NTP		Disinfectant
Hexachlorophene	61	MEIC, RC, TESS, NTP	Disinfectant	Amphetamine sulfate	55	MEIC, EDIT, RC (outlier), TESS, NTP	DEA	Pharmaceutical (stimulant)
Lindane	76	MEIC, EDIT, RC (outlier), EPA, NTP	Pesticide (insecticide)	Rotenone	60	RC, TESS, EPA, NTP		Pesticide (insecticide/ piscicide)

Table 3-2   Candidate Chemicals for the 3T3 and NHK NRU Test Methods Validation Study

Selected Chemicals				Alternate Chemicals				
GHS Category[1]/Chemical	Rodent Oral LD$_{50}$[2] (mg/kg)	Source[3]	Product/Use[4]	GHS Category[1]/Chemical	Rodent Oral LD$_{50}$[2] (mg/kg)	Source[3]	Notes[5]	Product/Use[4]
Cadmium II chloride	88	RC, TESS, GD, NTP	Consumer/ industrial products	Furfural	65[a]	NTP, HPV		Solvent, food additive
Verapamil HCl	108	MEIC, EDIT, RC (outlier), TESS. NTP	Pharmaceutical (antiarrhythmic)	p-Phenylenediamine	80	RC, GD, NTP, HPV		Dyeing
Haloperidol	128[a]	MEIC, TESS	Pharmaceutical (antipsychotic)	Chlorpyrifos	82[a]	TESS, EPA, NTP		Pesticide (insecticide)
Sodium oxalate	155	MEIC, EDIT, RC, TESS, NTP	Paints, cleaners	Dextropropoxyphene HCl	83	MEIC, RC (outlier), TESS		Pharmaceutical (analgesic)
Phenobarbital	163	MEIC, RC (outlier), TESS, NTP	Pharmaceutical (anticonvulsant)	Methadone	86[a]	MEIC,TESS, NTP	DEA	Pharmaceutical (analgesic)
Sodium I fluoride	180	MEIC, RC, TESS, EPA, NTP	Electroplating, fluoridation	Fipronil	92[a]	EPA		Pesticide (insecticide)
Caffeine	192	MEIC, RC (outlier), TESS, NTP, HPV	Pharmaceutical (stimulant), food additive	Pentobarbital	125	MEIC, RC TESS	DEA	Pharmaceutical (sedative)
Diquat dibromide	231	MEIC, RC, TESS	Pesticide (herbicide)	Bromoxynil (phenol)	190[a]	EPA		Pesticide (herbicide)
Cupric sulfate * 5 H2O	300	MEIC, RC, TESS, EPA, NTP	Pesticide (insecticide/ fungicide)	Diphenylhydantoin	199	MEIC, RC, TESS, NTP		Pharmaceutical (anticonvulsant)
				Metaldehyde	227[a]	TESS, EPA		Pesticide (molluscicide)
				Carbaryl	230	RC, EPA. NTP		Pesticide (insecticide)
300 < LD$_{50}$ ≤ 2000 mg/kg								
Amitriptyline HCl	319	MEIC, EDIT, RC, TESS	Pharmaceutical (antidepressant)	Ferrous sulfate	319	MEIC, RC, TESS		Food additive
Phenol	414	MEIC, RC, TESS, EPA, NTP, HPV	Disinfectant	Warfarin	324	MEIC, RC, TESS, EPA		Pharmaceutical (anticoagulant), pesticide
Propranolol HCl	470[b]	MEIC, RC, TESS, GD	Pharmaceutical (antiarrhythmic)	Disopyramide	333[a]	MEIC, TESS		Pharmaceutical (antiarrythmic)
Chloral hydrate	479	MEIC, RC, TESS, NTP	Pharmaceutical (sedative)	Barium II nitrate	355	MEIC, RC, TESS, NTP		Pyrotechnic
Glutethimide	600	MEIC, RC,	Pharmaceutical	Thioridazine HCl	358	MEIC, RC, TESS		Pharmaceutical

Table 3-2   Candidate Chemicals for the 3T3 and NHK NRU Test Methods Validation Study

Selected Chemicals				Alternate Chemicals				
GHS Category[1]/Chemical	Rodent Oral LD$_{50}$[2] (mg/kg)	Source[3]	Product/Use[4]	GHS Category[1]/Chemical	Rodent Oral LD$_{50}$[2] (mg/kg)	Source[3]	Notes[5]	Product/Use[4]
		TESS	(sedative)					(antipsychotic)
Atropine sulfate	623	MEIC, EDIT, RC, TESS	Pharmaceutical (antimuscarinic)	Methylphenidate	367[a]	NTP	DEA	Pharmaceutical (stimulant)
Valproic acid	1695[b]	RC, MEIC, TESS, NTP	Pharmaceutical (anticonvulsant)	Molinate	369[a]	EPA, NTP		Pesticide (herbicide)
Meprobamate	794[a]	MEIC, TESS	Pharmaceutical (antidepressant)	2,4-Dichlorophenoxy-acetic acid	369	MEIC, RC, TESS, EPA, NTP, HPV		Pesticide (herbicide)
Acetylsalicylic acid	1000	MEIC, EDIT, RC, TESS, NTP	Pharmaceutical (analgesic)	Orphenadrine HCl	425	MEIC, RC, NTP		Pharmaceutical (analgesic)
Lithium I sulfate	1187[b]	MEIC, RC, TESS, NTP (Cl salt)	Pharmaceutical (mood stabilizer)	Trichlorfon	451	RC, EPA, GD, NTP		Pesticide (insecticide)
Procainamide	1950[a]	MEIC, TESS	Pharmaceutical (antiarrythmic)	Quinidine sulfate	456	MEIC, RC, NTP (base)		Pharmaceutical (antiarrhythmic)
Carbamazepine	1957[a]	MEIC, TESS	Pharmaceutical (antiepileptic)	1,3-Dichloropropene	470[a]	TESS, EPA, NTP		Pesticide (nematocide)
				Theophylline	600[b]	MEIC, RC, TESS, NTP		Pharmaceutical (antiasthmatic)
				Isoniazid	650	MEIC, RC, TESS, NTP		Pharmaceutical (antibiotic)
				Diazepam	709	MEIC, EDIT, RC, TESS, NTP	DEA	Pharmaceutical (anxiolytic)
				Maprotiline	760[a]	MEIC, TESS		Pharmaceutical (antidepressant)
				Methyleugenol	810[a]	NTP		Food additive
				Diphenhydramine HCl	855	MEIC, RC, TESS, NTP		Pharmaceutical (antihistamine)
				Malathion	885	MEIC, EDIT, RC, TESS, EPA, NTP		Pesticide (insecticide)
				Salicylic acid	891	RC, TESS, GD, NTP, HPV		Pharmaceutical (analgesic)
				Chloroform	908	MEIC, RC, NTP, HPV	Volatile (BP=61°C)	Solvent
				Chloroquine diphosphate	970	MEIC, RC		Pharmaceutical (antimalarial))

Table 3-2   Candidate Chemicals for the 3T3 and NHK NRU Test Methods Validation Study

Selected Chemicals				Alternate Chemicals				
GHS Category[1]/Chemical	Rodent Oral LD$_{50}$[2] (mg/kg)	Source[3]	Product/Use[4]	GHS Category[1]/Chemical	Rodent Oral LD$_{50}$[2] (mg/kg)	Source[3]	Notes[5]	Product/Use[4]
				Ibuprofen	1009	RC, TESS, GD		Pharmaceutical (analgesic)
				Nalidixic acid	1349	RC, GD, NTP		Pharmaceutical (antibiotic)
				Dichloromethane	1597	MEIC, RC, TESS, NTP, HPV	Volatile (BP=40°C)	Solvent
				Antipyrene	1800	RC, GD		Pharmaceutical (analgesic)
2000 < LD$_{50}$ ≤ 5000 mg/kg								
Acetaminophen	2404	MEIC, EDIT, RC, TESS, NTP	Pharmaceutical (analgesic)					
Potassium I chloride	2602	MEIC, RC, TESS, NTP	Pharmaceutical (electrolyte), manufacturing					
Boric aid	2660[a]	TESS, EPA, NTP	Pesticide (insecticide)					
Carbon tetrachloride	2799	MEIC, RC, TESS, NTP, HPV	Solvent					
Dimethylformamide	2800	RC, GD, NTP, HPV	Solvent					
Sodium chloride	2998	MEIC, EDIT, RC, TESS, EPA, NTP	Pharmaceutical (electrolyte), food additive					
Citric Acid	3000[a]	EPA, NTP, HPV	Food additive					
Chloramphenicol	3393	MEIC, RC, NTP	Pharmaceutical (antibiotic)					
Lactic acid	3730	RC, NTP, HPV	Food additive					
Acetonitrile	3798	RC, NTP, HPV	Solvent					
Xylene	4300	MEIC, RC, TESS, NTP, HPV	Solvent					
Trichloroacetic acid	4999	RC, NTP	Fixative					
LD$_{50}$ > 5000 mg/kg								

Table 3-2   Candidate Chemicals for the 3T3 and NHK NRU Test Methods Validation Study

Selected Chemicals				Alternate Chemicals				
GHS Category[1]/Chemical	Rodent Oral LD$_{50}$[2] (mg/kg)	Source[3]	Product/Use[4]	GHS Category[1]/Chemical	Rodent Oral LD$_{50}$[2] (mg/kg)	Source[3]	Notes[5]	Product/Use[4]
2-Propanol	5843	MEIC, RC, TESS, EPA, NTP, HPV	Disinfectant					
Gibberellic acid	6305	RC, EPA, NTP	Plant growth regulator					
Propylparaben	6326[b]	RC (outlier), NTP	Food additive					
5-Aminosalicylic acid	7749[b]	RC (outlier), NTP	Pharmaceutical (antibiotic)					
Ethylene glycol	8567	MEIC, EDIT, RC, TESS, NTP, HPV	Antifreeze					
Diethyl phthalate	8602	RC (outlier), NTP, HPV	Plasticizer					
Sodium hypochlorite	8910[d]	TESS, NTP	Disinfectant					
1,1,1-Trichloroethane	10298	MEIC, RC, NTP, HPV	Solvent					
Dibutyl phthalate	11998	RC (outlier), NTP, HPV	Plasticizer					
Glycerol	12691	RC, GD, NTP, HPV	Solvent					
Methanol	13012	MEIC, EDIT, RC, TESS, NTP, HPV	Solvent					
Ethanol	14008	MEIC, RC (outlier), TESS, EPA, NTP, HPV	Solvent					

160   [1]GHS-Globally Harmonized System of Classification and Labelling of Chemicals for acute oral toxicity (UN 2005).
161   [2]LD$_{50}$ data are from the Registry of Cytotoxicity (Halle 1998) and are for rats, the preferred species for oral acute toxicity studies, unless otherwise noted. Data with decimal
162   places are rounded to the nearest one.
163   [3]Sources used to identify candidate chemicals: EDIT = Evaluation-guided Development of New *In vitro* Test Batteries; EPA = pesticides registered with the Environmental
164   Protection Agency; EHS = EPA's Extremely Hazardous Substance list; HPV = High Production Volume chemicals (i.e., those that are imported into or produced in the United
165   States in amounts $\geq$ 1,000,000 lbs/year; GD = *Guidance Document* (ICCVAM 2001b); MEIC = Multicentre Evaluation of In Vitro Cytotoxicity; NTP = National Toxicology
166   Program; RC = Registry of Cytotoxicity with chemicals classified as regression outliers shown in parentheses; TESS = Toxic Exposure Surveillance System (Litovitz et al. 2000).

167   [4]Product/use categories from Hazardous Substances Data Bank (NLM 2002) or Registry of Toxic Effects of Chemical Substances ([RTECS[®]], MDL Information Systems 2002).
168   Pharmaceutical uses from Gilman et al. (1985) or Thomson PDR[®] (2004).
169   [5]Only chemicals expected to be too volatile for the cytotoxicity assay system have "volatile" notations. BP = Boiling point. DEA (U.S. Drug Enforcement Agency) refers to
170   Schedule II controlled substances. Chemicals with no "DEA" notation are expected to be under less strict control.
171   [a]RTECS[®] (MDL Information Systems 2002).
172   [b]Mouse.

173
174
175
176
177
178
179
180

- five chemicals associated with the highest incidence of toxic exposures reported by U.S. poison control centers participating in the TESS (Litovitz et al. 2000): hypochlorite, acetaminophen, ethanol, diphenhydramine, and isopropanol. The five chemicals with the greatest incidence of toxic exposures among children were the same, except that oxalate replaced ethanol. Most of these chemicals were already identified as candidate chemicals due to their inclusion in the MEIC study. Since hypochlorite (sodium salt) and diphenhydramine, were not already included, they were added to the list of candidates

181
182
183
184
185
186
187

- 11 chemicals recommended in the *Guidance Document* (ICCVAM 2001b) for qualifying *in vitro* cytotoxicity assays for the prediction of starting doses using the RC regression. These chemicals were recommended because the IC_{50} and LD_{50} data for these chemicals fit the RC regression line extremely well. These chemicals were sodium dichromate dihydrate, cadmium chloride, p-phenylenediamine, DL-propranolol HCl, trichlorfon, ibuprofen, nalidixic acid, salicylic acid, antipyrene, dimethylformamide, and glycerol

188

- 16 chemicals from the NTP

189
190
191
192
193
194
195
196
197

  o furfural, methyleugenol, and methylphenidate, scheduled for testing by the NTP National Center for Toxicogenomics (NCT) (G. Boorman, personal communication 2001), were added. Acetaminophen, another hepatotoxin to be tested by the NCT, was already a candidate chemical because it was included in the MEIC study. Chromium (VI), recommended by the NTP for consideration due to the potential for human exposure via drinking water (NTP 2002) was represented in the list of candidate chemicals by sodium dichromate dihydrate, which was also recommended in the *Guidance Document* (ICCVAM 2001b)

198
199
200

  o dibutyl phthalate, 5-aminosalicylic acid, propylparaben, gibberellic acid, and diethyl phthalate were added to increase the number of chemicals with LD_{50} values > 5000 mg/kg

201
202

  o trichloroacetic acid was added to increase the number of chemicals in the $2000 < LD_{50} \leq 5000$ mg/kg category

203   o   sodium selenate was added to increase the number of chemicals in the LD_{50}

204   ≤ 5 mg/kg category to 12

205   o   six chemicals that were also on the HPV list were added. Lactic acid, citric

206   acid, and acetonitrile were added to increase the number of chemicals in the

207             $2000 < LD_{50} \leq 5000$ mg/kg category. Tert-butylamine, 2,4-dinitrophenol,

208   and acrolein were added to increase the number of chemicals in the $5 < LD_{50}$

209   ≤ 50 mg/kg category

210   •   eight additional RC chemicals in the $LD_{50} \leq 5$ mg/kg category. These were:

211   triethylenemelamine, busulfan, disulfoton, parathion, aminopterin,

212   phenylthiourea, epinephrine bitartrate, and aflatoxin B1

213

214   The goal to identify more than 12 candidate chemicals for each toxicity category was

215   unrealized for three toxicity categories. The most toxic category ($LD_{50} \leq 5$ mg/kg), and least

216   toxic categories ($2000 < LD_{50} \leq 5000$ mg/kg, $LD_{50} > 5000$ mg/kg), contained only 12

217   candidate chemicals each. The intermediate toxicity categories ($50 < LD_{50} \leq 300$ mg/kg, $>$

218   $300 < LD_{50} \leq 2000$ mg/kg), however, contained two to three times the minimum number of

219   candidate chemicals.

220

221   3.1.3   Selection of Reference Substances for Testing

222   Using the candidate chemical database, 72 reference substances (12 unclassified chemicals

223   and 12 chemicals from each the five GHS acute oral toxicity hazard categories) were selected

224   for use in the NICEATM/ECVAM validation study. The criteria for prioritizing the

225   candidate chemicals were:

226   •   the availability of rodent acute oral toxicity data (e.g., RC, RTECS[®])

227   •   the availability of human acute oral toxicity data and/or relevance for human

228   exposure (e.g., MEIC, EDIT, TESS, NTP)

229   •   the lack of excessive volatility as estimated by SMT chemical consultants.

230   Since the cells are exposed for 48 hours while incubated at 37°C in 96-well

231   plates, volatilization from wells with high reference substance concentrations

232   would reduce the extent of cytotoxicity and potentially contaminate other wells

233   in close proximity

234   • the lack of U.S. Drug Enforcement Agency (DEA) controls. Excluding

235   chemicals that are listed in DEA Schedules I and II from consideration obviates

236   the requirement for U.S. laboratories to obtain a DEA license and adhere to

237   strict chemical storage and control procedures

238   • practical considerations such as cost and disposal issues

239

240   If more than twelve candidate chemicals in a GHS category met the above criteria, then

241   selection was based on two further considerations. One consideration was the distribution of

242   chemical toxicities within each toxicity category (i.e., the goal was to select chemicals that

243   represented the entire range of toxicity within each category). Another consideration, which

244   applied only to candidate chemicals selected from the RC database, was the fit of the

245   chemical to the RC regression. Chemicals with the best fit to the RC regression were

246   preferentially selected to prevent the entire set of reference substances from having

247   proportionally more "outlier" substances (i.e., greater than one-half log from the RC

248   regression) than the entire RC database.

249

250   The final list of selected reference substances is provided by GHS acute oral toxicity

251   category on the left side of **Table 3-2.**

252

253   **3.2   Rationale for the Number of Reference Substances Selected**

254

255   Seventy-two reference substances were used to evaluate the ability of the 3T3 and NHK

256   NRU test methods to estimate the acute oral LD_{50} and thus the starting dose for *in vivo* acute

257   oral toxicity tests. The SMT determined the number of reference substances for testing by

258   first using the GHS classification scheme for acute oral toxicity (UN 2005) to assure that the

259   candidate chemicals covered the complete range of toxicity, (see **Table 3-1**) then deciding

260   how many chemicals would be tested per category. To adequately cover the range of toxicity

261   within each of the six toxicity groups, the SMT decided to test 12 chemicals per group.

262   Seventy-two reference substances (12 substances/group with six groups) were deemed

263   adequate by the SMT, the ICCVAM Acute Toxicity Working Group (ATWG), ICCVAM,

264   and ECVAM.

265

266   The total number of reference substances was comparable to the number used in other

267   contemporary *in vitro* test method multilaboratory validation studies. The European

268   Cosmetic, Toiletry, and Perfumery Association evaluation of multiple alternatives to the

269   Draize eye irritation test used 55 reference substances (Brantom et al. 1997). ECVAM's

270   evaluations of *in vitro* dermal corrosivity test methods (Fentem et al. 1998) and *in vitro*

271   dermal irritation assays (Botham 2004) used 60 reference substances.

272

273   **3.3   Characteristics of the Selected Reference Substances**

274

275   The physical/chemical and toxicological information in **Appendix F** may be useful for

276   characterizing the performance of the *in vitro* NRU assays for various chemical types.

277   **Appendix F-1** lists the selected reference substances in alphabetical order with information

278   on the CASRN, purity, supplier, pH, and concentrations tested in the *in vitro* NRU

279   cytotoxicity assays. **Appendix F-2** also provides the reference substances in alphabetical

280   order, but with the available information on molecular weight, chemical class, water

281   solubility, acid/base dissociation constant (pK), boiling point, lipid solubility (log K_{ow}),

282   major toxic effects, ability to pass the blood:brain barrier, metabolic activation/inactivation,

283   and mechanism of lethality. The remainder of **Section 3.3** summarizes selected

284   characteristics of the reference substances.

285

286   3.3.1   Source Databases Represented by the Selected Reference Substances

287   The primary sources of chemicals, which reflect the level of societal interest, were well

288   represented in the final list of reference substances. **Table 3-3** shows the distribution of

289   reference substances by GHS category from the MEIC, EDIT, TESS, NTP, and HPV lists.

290   Forty-two (58%) of the 72 selected chemicals were MEIC chemicals (17 of the 42 MEIC

291   chemicals [40%] were EDIT chemicals), 46 (64%) chemicals were involved in human

292   poisonings report by TESS, 51 (71%) chemicals have been evaluated by the NTP, and 18

293   (25%) chemicals are listed in the EPA's HPV Challenge Program. Some chemicals were

294   found in more than one source.

295

296   The other major source of chemicals was the RC. As shown in **Table 3-4**, 58 (81%) of the

297   selected chemicals were included in the RC. Since one of the regression formulas evaluated

298   in the NICEATM/ECVAM validation study was the RC regression, the fit of the RC

299   chemicals to the regression was relevant (Halle 1998). Halle (1998) defined outliers as those

300   chemicals with log IC_{50}-log LD_{50} points that were > 0.699 (i.e., log 5) from the RC

301   regression. For each toxicity category, **Table 3-4** shows the number of RC outliers selected

302   for testing and the corresponding number of outliers in the RC. Although the percentage of

303   outliers for the selected chemicals in several GHS categories is similar to the RC, the total

304   percentage of RC outliers in the set of reference substances (i.e., 38% [22/58]) is greater than

305   the total percentage of outliers in the RC (i.e., 27% [95/347]). For the reference substances,

306   the RC prediction model underpredicted toxicity (i.e., actual LD_{50} is lower than predicted) for

307   17 outliers and overpredicted toxicity (i.e., actual LD_{50} is higher than predicted) for five

308   outliers. **Figure 3-1** shows the 58 RC chemicals selected for testing with the remaining 289

309   RC data points and the RC regression. In the figure, the outliers are those points outside the

310   RC prediction interval. The 17 outlier chemicals for which toxicity is underpredicted are

311   below the lower prediction interval and the five outliers for which toxicity is overpredicted

312   are above the upper prediction interval.

313

314   **Table 3-3   Distribution of Candidate Chemicals and Reference Substances by Source[1] and Toxicity Category**

GHS[2] Category (LD$_{50}$ in mg/kg)	Reference Substances/ Candidate Chemicals	MEIC Reference/ MEIC Candidates	EDIT Reference/ EDIT Candidates	TESS Reference/ TESS Candidates	NTP Reference/ NTP Candidates	HPV Reference/ HPV Candidates
LD$_{50}$ ≤ 5	12/13	2/2	1/1	3/3	5/5	0/0
5 < LD$_{50}$ ≤ 50	12/15	6/6	5/5	9/10	8/11	2/5
50 < LD$_{50}$ ≤ 300	12/26	11/17	4/5	11/19	9/18	1/3
300 < LD$_{50}$ ≤ 2000	12/38	12/29	3/5	12/27	5/23	1/5
2000 < LD$_{50}$ ≤ 5000	12/12	6/6	2/2	6/6	12/12	6/6
LD$_{50}$ > 5000	12/12	5/5	2/2	5/5	12/12	8/8
Total	72/116	42/65	17/20	46/70	51/81	18/27

315   [1]Substances may be represented in more than one source (see **Table 3-2**).
316   [2]GHS = Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005).
317   MEIC = Multicentre Evaluation of In Vitro Cytotoxicity**;** EDIT= Evaluation-Guided Development of *In vitro* Tests; TESS =Toxic Exposure Surveillance
318   System; NTP = U.S. National Toxicology Program; HPV = EPA High Production Volume program.
319
320
321

321   **Table 3-4**   **Selected Chemicals: Distribution of Registry of Cytotoxicity (RC) Chemicals and RC Outliers[1] by Toxicity**

322   **Category**

GHS[2] Category (LD$_{50}$ in mg/kg)	RC Outliers/ Total Chemicals	Candidate and Selected Chemicals		
		Candidate Chemicals	RC Reference / RC Candidates	RC Reference Outliers/ RC Reference Chemicals
LD$_{50}$ ≤ 5	10/11 (91%)	13	9/10	8/9 (89%)
5 < LD$_{50}$ ≤ 50	15/26 (58%)	15	8/10	4/8 (50%)
50 < LD$_{50}$ ≤ 300	24/70 (34%)	26	11/18	5/11 (45%)
300 < LD$_{50}$ ≤ 2000	14/139 (10%)	38	9/29	0/9 (0%)
2000 < LD$_{50}$ ≤ 5000	12/57 (21%)	12	10/10	0/10 (0%)
LD$_{50}$ > 5000	20/44 (45%)	12	11/11	5/11 (45%)
Total	95/347 (27%)	116	58/88	22/58 (38%)

323   [1]Chemicals falling outside the log 5 (i.e., > ± 0.699) prediction interval for the RC regression (Halle 1998).
324   [2]GHS: Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005).
325
326

327   **Figure 3-1   The Fifty-Eight (58) Selected Registry of Cytotoxicity (RC) Chemicals on**
328   **the RC Regression**

329
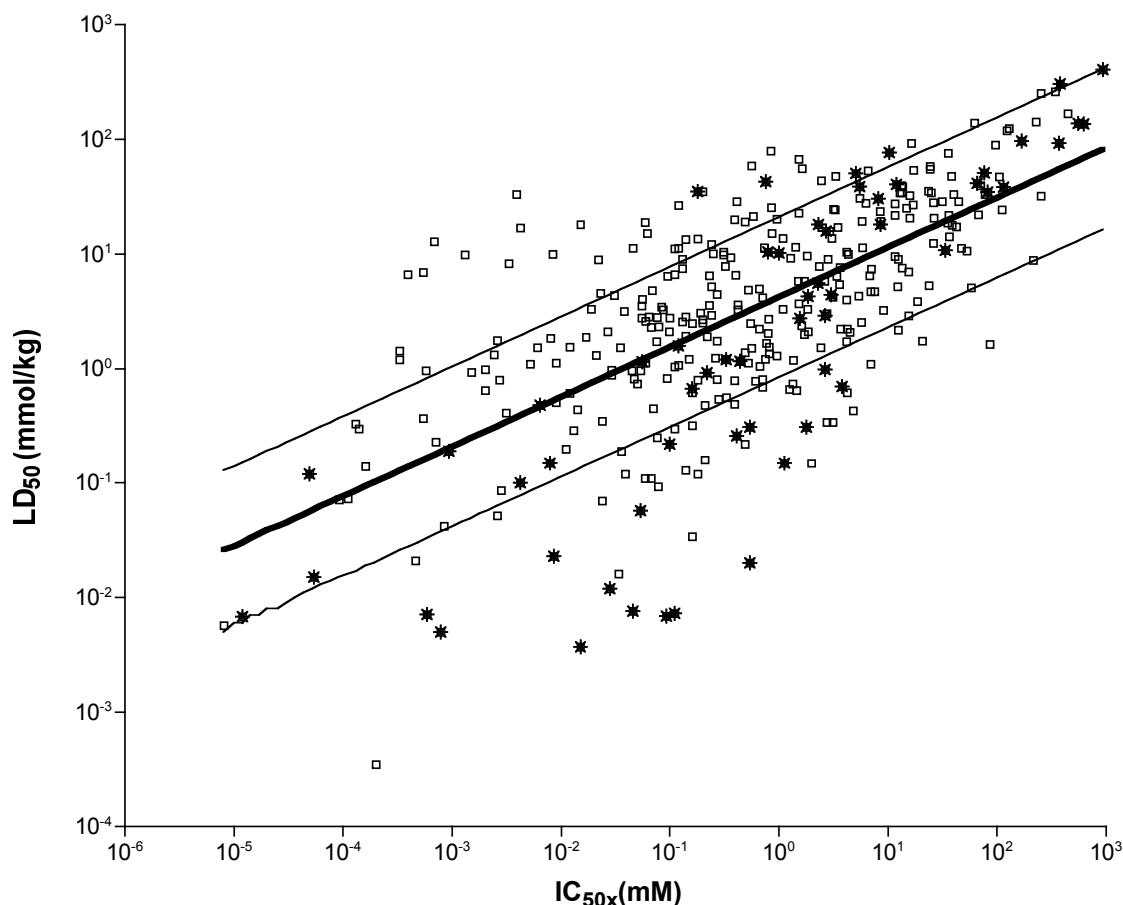
330   The 58 RC chemicals tested in the NICEATM/ECVAM validation study are shown by *. The RC
331   regression, log (LD$_{50}$) = 0.435 x log (IC$_{50x}$) + 0.625, is shown by the bold line. The lighter lines show
332   the ± log 5 (i.e., ± 0.699) prediction interval (Halle 1998). The remaining 289 RC data points are
333   shown by the open boxes.

334

335

336   3.3.2   Chemical Classes Represented by the Selected Reference Substances

337   Medical subject headings (MeSH) from the NLM were used to determine chemical class for

338   the selected chemicals. Of the 72 reference substances, 55 (76%) were organic compounds

339   and 17 (24%) were inorganic compounds. The most commonly represented classes of

340   organic compounds were heterocyclic compounds (14/55, 26%), carboxylic acids (12/55,

341   22%), and alcohols (10/55, 18%). **Table 3-5** shows the distribution of the selected chemicals

342   among the GHS toxicity categories. The 14 heterocyclic compounds were rather evenly

343   distributed among the first four GHS toxicity categories for $LD_{50} \leq 2000$ mg/kg with the

344   majority of the heterocyclics (11/14) in the categories for $LD_{50} < 300$ mg/kg. The majority

345   of the carboxylic acids (10/12) and alcohols (8/10) had $LD_{50} > 300$ mg/kg. The majority of

346   the inorganic compounds (12/17) had $LD_{50} < 300$ mg/kg.

347

348   3.3.3   <u>Product/Use Classes Represented by the Selected Reference Substances</u>

349   Product and use information for the selected chemicals was obtained from HSDB (NLM

350   2002) or RTECS[®] (MDL Information Systems 2002). Since more than one use was reported

351   for some chemicals, the number of assigned uses (77) is greater than the number of selected

352   chemicals. **Table 3-6** shows the distribution of products and uses of the selected chemicals

353   among the GHS toxicity categories. Pharmaceutical (27/77; 35%) and pesticide (17/77;

354   22%) uses were observed most frequently. The toxicity category for $300 < LD_{50} \leq 2000$

355   mg/kg had the highest number of chemicals with pharmaceutical uses. Every toxicity

356   category except for $LD_{50} > 5000$ mg/kg had at least four chemicals with pharmaceutical uses.

357   The majority of chemicals (16/17; 94%) with pesticide uses had $LD_{50} < 300$ mg/kg. The next

358   most frequent uses for the selected chemicals were solvents (8/77; 10%) and food additives

359   (5/77; 6%). The toxicity categories for $LD_{50} > 2000$ mg/kg contained most of the chemicals

360   with solvent (8/8; 100%) and food additive (4/5; 80%) uses.

361

362   3.3.4   <u>Toxicological Characteristics of the Selected Reference Substances</u>

363   *Corrosivity*

364   During the chemical selection process, the intent of the SMT was prioritize chemicals with

365   low corrosivity because guidelines for acute systemic toxicity testing indicate that corrosive

366   or severely irritating chemicals need not be tested (OECD 2001a, c, d). The UN and U.S.

367   Department of Transportation Packing Group (DOT PG) classification system was used to

368   classify the corrosivity hazard associated with the candidate chemicals. However, after

369   chemical selection was completed and testing had begun, the SMT discovered that the PG

370   classification system is also based on hazards other than corrosivity (e.g., dermal and

371   inhalation toxicity, flammability, etc.). Thus, the selected chemicals were not actually

372   prioritized by corrosivity. Subsequent information on the corrosivity of the selected

373   chemicals was obtained from HSDB (NLM 2004) and the Material Safety Data Sheets

374   (MSDS) provided with the purchased reference substances. Seven substances had corrosive

375   notations. The MSDSs for lactic acid, sodium hypochlorite, sodium oxalate, and

376   trichloroacetic acid indicated that these chemicals should carry a corrosive label. Chloral

377   hydrate, mercury II chloride, and potassium cyanide were noted to be corrosive to eyes or

378   skin by their HSDB files.

379

Table 3-5   Distribution of Chemical Class for the 72 Reference Substances by Toxicity Category

Chemical Class[2]	GHS Acute Oral Toxicity Category[1] (mg/kg)						Total
	≤ 5	> 5 - ≤ 50	> 50 - ≤ 300	> 300 - ≤ 2000	> 2000 - ≤ 5000	> 5000	
Organic							
Heterocyclic compound	4	3	4	3	0	0	14
Carboxylic acid	1	0	1	3	3	4	12
Alcohol	2	0	0	2	1	5	10
Amide	0	0	0	1	2	0	3
Halogenated hydrocarbon	0	0	1	0	1	1	3
Cyclic hydrocarbon	0	0	1	0	1	0	2
Hydrocarbon	0	1	0	0	0	1	2
Organophosphorous compound	2	1	0	0	0	0	3
Polycyclic compound	0	1	0	1	0	0	2
Amine	0	0	1	0	0	0	1
Nitrile	0	0	0	0	1	0	1
Organometallic compound	0	1	0	0	0	0	1
Phenol	0	0	0	1	0	0	1
Total	9	7	8	11	9	11	55
Inorganic							
Arsenical	0	2	0	0	0	0	2
Sulfur compound	1	0	1	0	0	0	2
Boron compound	0	0	0	0	1	0	1
Cadmium compound	0	0	1	0	0	0	1
Ketone	0	0	1	0	0	0	1
Lithium compound	0	0	0	1	0	0	1
Mercury compound	1	0	0	0	0	0	1
Metal	0	1	0	0	0	0	1

Table 3-5   Distribution of Chemical Class for the 72 Reference Substances by Toxicity Category

Chemical Class[2]	GHS Acute Oral Toxicity Category[1] (mg/kg)						Total
	≤ 5	> 5 - ≤ 50	> 50 - ≤ 300	> 300 - ≤ 2000	> 2000 - ≤ 5000	> 5000	
Potassium, chlorine compound	0	0	0	0	1	0	1
Potassium, nitrogen compound	0	1	0	0	0	0	1
Sodium, chlorine compound	0	0	0	0	1	0	1
Sodium, chromium compound	0	1	0	0	0	0	1
Sodium, fluorine compound	0	0	1	0	0	0	1
Sodium, oxygen, chlorine compound	0	0	0	0	0	1	1
Sodium, selenium compound	1	0	0	0	0	0	1
Total	3	5	4	1	3	1	17

380 [1]GHS: Globally Harmonized System of Classification and Labelling of Chemicals based on oral LD_{50} (UN 2005).
381 ≤ 5: $LD_{50} \leq 5$ mg/kg
382 > 5 - ≤ 50: $5 < LD_{50} \leq 50$ mg/kg
383 > 50 - ≤ 300: $50 < LD_{50} \leq 300$ mg/kg
384 > 300 - ≤ 2000: $300 < LD_{50} \leq 2000$ mg/kg
385 > 2000 - ≤ 5000: $2000 < LD_{50} \leq 5000$ mg/kg
386 > 5000: $LD_{50} > 5000$ mg/kg
387 [2]Based on the Medical Subject Heading [MeSH] index (NLM 2005).

388

388   **Table 3-6   Distribution of Product/Use[1] Class for the 72 Reference Substances by Toxicity Category**

Product/Use Class[1]	GHS Acute Oral Toxicity Category[2] (mg/kg)						Total
	≤ 5	> 5 - ≤ 50	> 50 - ≤ 300	> 300 - ≤ 2000	> 2000 - ≤ 5000	> 5000	
Antibiotic/fungicide	1	0	0	0	0	0	1
Antifreeze	0	0	0	0	0	1	1
Consumer/industrial products	0	0	1	0	0	0	1
Disinfectant	0	0	1	1	0	2	4
Electroplating	0	2	0	0	0	0	2
Fluoridation	0	0	1	0	0	0	1
Feed additive	1	0	0	0	0	0	1
Fixative	0	0	0	0	1	0	1
Food additive	0	0	1	0	3	1	5
Manufacturing	1	0	0	0	1	0	2
Oxidizing agent	0	1	0	0	0	0	1
Paints, cleaners	0	0	1	0	0	0	1
Pesticide	5	7	4	0	1	0	17
Pharmaceutical	4	3	4	11	4	1	27
Plant growth regulator	0	0	0	0	0	1	1
Plasticizer	0	0	0	0	0	2	2
Preservative	1	0	0	0	0	0	1
Solvent	0	0	0	0	4	4	8

389   [1]Product/use categories from Hazardous Substances Data Bank (NLM 2002) or Registry of Toxic Effects of Chemical Substances ([RTECS®], MDL Information
390   Systems 2002). Some chemicals are counted more than once due to multiple uses.
391   ≤ 5: $LD_{50} \leq 5$ mg/kg
392   > 5 - ≤ 50: $5 < LD_{50} \leq 50$ mg/kg
393   > 50 - ≤ 300: $50 < LD_{50} \leq 300$ mg/kg
394   > 300 - ≤ 2000: $300 < LD_{50} \leq 2000$ mg/kg
395   > 2000 - ≤ 5000: $2000 < LD_{50} \leq 5000$ mg/kg
396   > 5000: $LD_{50} > 5000$ mg/kg

397   [2]GHS: Globally Harmonized System of Classification and Labelling of Chemicals based on oral LD_{50} (UN 2005).

398

399   *Toxicity Targets*

400   As shown in **Appendix F**, the most common toxicological effects were neurological (40

401   reference substances); 26 reference substances cause central nervous system (CNS)

402   depression, seven reference substances produce CNS stimulation, four reference substances

403   produce other CNS affects such as encephalopathy, and three reference substances attack the

404   peripheral nervous system. Other common toxicity targets include the liver (17 reference

405   substances), kidney (15 reference substances), and cardiovascular system (10 reference

406   substances). No target organ information was available for gibberellic acid. Among the 72

407   reference substances, 27 had multiple toxicity targets.

408

409   *Metabolism*

410   **Table 3-7** shows the 22 reference substances that are known or expected to produce

411   active/toxic metabolites *in vivo*. In contrast, dichlorvos, fenpropathrin, meprobamate,

412   phenylthiourea, and sodium dichromate are known to be rapidly inactivated by metabolism *in*

413   *vivo* to less toxic compounds. Because the NHK and 3T3 cells have little (see Babich 1991)

414   or no metabolic capability, respectively, metabolites of these compound would be

415   unavailable *in vitro*. See **Appendix F-2** for more information on the metabolism of the

416   selected chemicals.

417

418   **Table 3-7   Reference Substances Metabolized to Active Metabolites**

Known to Have Active Metabolites				Active Metabolites Expected
Acetaminophen	Carbamazepine	Digoxin	Methanol	Carbon tetrachloride
Acetonitrile	Chloral hydrate	Disulfoton	Parathion	Triethylenemelamine
Acetylsalicylic acid	Cycloheximide	Ethanol	Procainamide HCl	Valproic acid
Amitriptyline HCl	Dibutyl phthalate	Ethylene glycol	Verapamil HCl	
Busulfan	Diethyl phthalate	Glutethimide		

419

420

421   3.3.5   Selection of Reference Substances for Testing in Validation Study Phases Ib and II

422   Based on the *Guidance Document* (ICCVAM 2001b) recommendation that 10-20 chemicals

423   be tested to qualify candidate *in vitro* cytotoxicity tests for determining starting doses for

424   acute oral systemic toxicity assays, 12 reference substances were chosen from the 72

425   reference substances for testing in Phases Ib and II of the validation study (see **Table 3-8**).

426   The criteria for choosing these reference substances, in order of importance, were:

427   • two reference substances must be included from each of the five GHS toxicity

428   categories and the unclassified category

429   • the log LD_{50} (mmol/kg) must be within 0.699 of the RC regression (i.e., within

430   the RC prediction interval). The *Guidance Document* (ICCVAM 2001b)

431   recommends that reference substances for evaluating a cytotoxicity test to use

432   with the RC regression fit the regression as closely as possible

433   • MEIC chemicals must be included. Cytotoxicity data from these phases (and

434   Phase III of this study) and the available human toxicity information for the

435   MEIC chemicals could be used to build a prediction model for estimating

436   human lethal blood concentrations. Phase Ib reference substances arsenic

437   trioxide and ethylene glycol are EDIT chemicals

438

439   If more than two chemicals in a GHS category met the above criteria, reference substances

440   were selected so that the LD_{50} was as close to the RC prediction as possible and/or to

441   represent the range of toxicity in each GHS category.

442

443   Only nine reference substances of the 72 selected reference substances fit all three criteria.

444   One reference substance was not within the RC acceptance interval. For the most toxic

445   category (i.e., $LD_{50} \leq 5$ mg/kg), only one RC chemical, aminopterin, was within 0.699 of the

446   RC regression. Sodium selenate, whose fit to the RC regression was unknown and had not

447   been tested in the MEIC study, was included in this toxicity category. In addition, neither of

448   the two reference substances chosen for the $LD_{50} \leq 5$ mg/kg category, aminopterin and

449   sodium selenate, were MEIC chemicals.

450

451

452

453

454

455

455   **Table 3-8   Reference Substances Tested in Phases Ib and II**

Reference Substances	CASRN	RC Reference No.	MEIC Reference No.	Rodent Oral LD$_{50}$[1] (mg/kg)	Observed – Predicted log LD$_{50}$[2]
LD$_{50}$ ≤ 5 mg/kg					
Aminopterin	54-62-6	3	NA	3	-0.652
Sodium selenate	13410-01-0	NA	NA	1.6[3]	NA
5 < LD$_{50}$ ≤ 50 mg/kg					
Colchicine	64-86-8	6	60	6[4]	-0.593
Arsenic III trioxide	1327-53-3	153	26	20	-0.591
50 < LD$_{50}$ ≤ 300 mg/kg					
Cadmium II chloride	10108-64-2	81	NA	88	-0.336
Sodium I fluoride	7681-49-4	106	14	180	-0.109
300 < LD$_{50}$ ≤ 2000 mg/kg					
DL-Propranolol HCl	350-60-90	54	23	470[4]	-0.023
Lithium I carbonate	544-13-2	327[4]	20	1187[4,5]	-0.256[4]
2000 < LD$_{50}$ ≤ 5000 mg/kg					
Potassium I chloride	7447-40-7	346	50	2602	0.085
Chloramphenicol	56-75-7	91	45	3393	0.441
LD$_{50}$ > 5000 mg/kg					
2-Propanol	67-63-0	128	10	5843	0.396
Ethylene glycol	107-21-1	360	7	8567	0.321

456   [1]From the RC (Halle 1998) unless otherwise indicated. Data for rats unless otherwise indicated.
457   [2]Available only for chemicals included in the RC. This figure characterizes the log LD$_{50}$ deviation from the RC
458   regression. Outliers are > ± 0.699 from the regression line.
459   [3]RTECS® (MDL Information Systems 2002).
460   [4]Mouse data.
461   [5]Data for lithium sulfate.
462   Abbreviations: CASRN = Chemical Abstracts Service Registry Number; RC = Registry of Cytotoxicity; MEIC
463   = Multicentre Evaluation of *In Vitro* Cytotoxicity; NA – not applicable; chemical not included in the RC and/or
464   MEIC studies.
465

466

467   3.3.6   Unsuitable and Challenging Reference Substances

468   Several reference substances could not be adequately tested for cytotoxicity in either or both

469   of the 3T3 or NHK NRU test methods. Under the conditions of the NRU cytotoxicity test,

470   the following reference substances did not produce sufficient toxicity at soluble

471   concentrations for calculation of an IC$_{50}$ at the highest concentrations that could be tested:

472   • carbon tetrachloride in either the 3T3 or NHK test method in all three

473   laboratories

474   • xylene in either test method in two laboratories

475　　　　　• methanol in the 3T3 test method in all three laboratories and in the NHK test
476　　　　　　method in two laboratories

477　　　　　• lithium carbonate in the 3T3 test method in two laboratories

478　　　　　• 1,1,1-trichloroethane in the NHK test method in two laboratories

479　　　　　• valproic acid in the 3T3 test method in one laboratory

480

481　Other reference substances were difficult to test, but three acceptable tests were obtained

482　after a number of trials.

483　　　　　• Acetonitrile and 2-propanol were so volatile and nontoxic that, even with the
484　　　　　　use of film plate sealers, one to seven tests failed at each laboratory. Tests with
485　　　　　　these two reference substances often failed the VC and data points criteria.

486　　　　　• Disulfoton failed at least one test in both test methods in two laboratories due to
487　　　　　　inadequate toxicity and solubility.

488　　　　　• Dibutyl phthalate failed one 3T3 test at one laboratory and one NHK test at one
489　　　　　　laboratory due to inadequate toxicity and solubility.

490　　　　　• Lindane failed one 3T3 test due to inadequate toxicity and solubility and one
491　　　　　　3T3 test due to volatility.

492　　　　　• Parathion failed one test due to inadequate toxicity and solubility in both the
493　　　　　　3T3 and NHK test methods and one NHK test due to volatility.

494　　　　　• Diethyl phthalate failed one NHK test due to volatility.

495　　　　　• Digoxin, gibberellic acid, and strychnine failed at least one 3T3 test in more
496　　　　　　than one laboratory due to inadequate toxicity and solubility.

497

498　**3.4　　　Reference Substance Procurement, Coding, and Distribution**

499

500　Reference substances were purchased from the suppliers in the purities indicated in

501　**Appendix F** and distributed by BioReliance Corporation (Rockville, MD). BioReliance also

502　collected information from the suppliers on the analytical purity, composition, and stability

503　of the reference substances. BioReliance tested the reference substances for solubility,

504　packaged them into 4 g aliquots for shipment to the cytotoxicity testing laboratories, and

505　archived two additional samples. All reference substances were randomly coded to conceal

506   the identities from the cytotoxicity testing laboratories. Each reference substance had a code

507   unique for each testing facility. About 100 g of the positive control, SLS, was distributed to

508   each laboratory and one additional sample was archived.

509

510   Reference substances were packaged to minimize damage during transit and shipped under

511   appropriate storage conditions and according to proper regulatory transportation procedures.

512   Testing facilities were notified upon shipment in order to prepare for receipt. With the

513   exception of the positive control shipment, which was shipped directly to the Study

514   Directors, the reference substances were shipped to the test facility Safety Officers.

515   Reference substances shipments were accompanied by a sealed information packet

516   containing the appropriate health and safety procedures for use (i.e., MSDS or equivalent

517   documentation with information regarding the proper protection for handling, procedures for

518   dealing with accidental ingestion or contact with skin or eyes, and procedures for containing

519   and recovering spills) and a disclosure key for identifying reference substances by code.

520   Also provided was a data sheet giving a minimum of essential information for each reference

521   substance, including color, odor, physical state, weight or volume of sample, specific density

522   for liquid reference substances, and storage instructions. The shipment directed the Safety

523   Officer to:

524   • notify BioReliance and the SMT upon receipt of reference substances

525   • retain the health and safety package and provide the reference substances and

526   chemical data sheets to the Study Director without revealing the identities of the

527   reference substances

528   • notify the SMT if test facility personnel open the health and safety packet at any

529   time during the study

530   • return the unopened health and safety package to BioReliance after testing is

531   complete

532

533   *Exceptions*

534   The Safety Officer for ECBC required the information on reference substance codes before

535   the substances were shipped to the Safety Office to satisfy the facility's environmental

536   procedures and requirements. The reference substance codes were stored in a classified safe

537   located in the Safety Office, which was in a building separate from the cytotoxicity testing

538   laboratory. Cytotoxicity testing personnel had no access to the reference substance codes.

539   The ECBC Safety Officer opened the sealed health and safety packets for lithium carbonate

540   and ethanol upon receipt of those substances because the code information for these

541   substances was not included in the list originally provided. ECBC cytotoxicity testing

542   personnel never had access to the reference substance codes.

543

544   **3.5   Reference Substances Recommended by the** *Guidance Document* (ICCVAM

545   2001b)

546

547   The *Guidance Document* method for evaluating basal cytotoxicity assays for use in

548   predicting starting doses for acute oral toxicity assays provides the existing performance

549   standard (ICCVAM 2001b) for the 3T3 and NHK NRU test methods. The *Guidance*

550   *Document* specifically recommends testing the following 11 chemicals to qualify candidate

551   basal cytotoxicity assays: sodium dichromate dihydrate, cadmium chloride, p-

552   phenylenediamine, DL-propranolol HCl, trichlorfon, ibuprofen, nalidixic acid, salicylic acid,

553   antipyrene, dimethylformamide, and glycerol (ICCVAM 2001b). Although the 11 reference

554   chemicals recommended in the *Guidance Document* were considered as candidates for

555   testing in the NICEATM/ECVAM validation study (see **Section 3.1.2**), only sodium

556   dichromate dihydrate, cadmium chloride, DL-propranolol HCl, dimethylformamide, and

557   glycerol were chosen for testing after the candidate chemicals were prioritized as described

558   in **Section 3.1.3**. The other seven were excluded based on the criterion hierarchy used to

559   determine the selected chemicals (e.g., were not MEIC chemicals, not identified as high

560   exposure risk in TESS)

561

562   **3.6   Summary**

563

564   Seventy-two reference substances were selected for testing in the NICEATM/ECVAM

565   validation study. The reference substances were selected to represent: (1) the complete range

566   of *in vivo* acute oral toxicity ranges (in terms of LD_{50} values); (2) the types of substances

567   regulated by various regulatory authorities; and (3) those with human toxicity data and/or

568   human exposure potential. To assure the complete range of toxicity was covered, the

569   Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005) was

570   used to select 12 chemicals for each acute oral toxicity category and 12 unclassified

571   chemicals. The set of selected reference substances had the following characteristics:

572   • 38% (27/72) of the substances had pharmaceutical uses, 21% (15/72) had

573   pesticide uses, 11% (8/72) had solvent uses, and 7% (5/72) had food additive

574   uses. The remaining substances were used for a variety of manufacturing and

575   consumer products

576   • relevance of the substances to human exposures was indicated by the fact that

577   58% (42/72) were included in the MEIC study, 24% (17/72) were included in

578   the EDIT program, 64% (46/72) had human exposures reported by TESS, 71%

579   (51/72) had been evaluated by NTP, and 25% (18/72) were included in EPA's

580   HPV list

581   • 81% (58/72) of the substances were also included in the RC and 38% (22/58) of

582   these were outliers with respect to the RC regression

583   • 76% (55/72) were organic compounds and 24% (17/72) were inorganic

584   compounds. The most commonly represented classes of organic compounds

585   were heterocyclic compounds (26%, 14/55), carboxylic acids (22%, 12/55), and

586   alcohols (18%, 10/55)

587   • 19 substances (26%, 19/72,) were known to have active metabolites and three

588   additional substances were expected to have active metabolites

589   • many of the selected chemicals had multiple target organs. The most common

590   effects were neurological (40 chemicals), liver (17 chemicals), kidney (15

591   chemicals), and cardiovascular (10 chemicals). No target organ information

592   was available for one chemical (gibberellic acid)

593

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608   ***[This Page Intentionally Left Blank]***

609

610

17
18
19
20
21
22
23
24
25
26
27
28
29
30
31   ***[This Page Intentionally Left Blank]***
32

32   **4.0**   *IN VIVO* **RODENT TOXICITY REFERENCE VALUES USED TO ASSESS**

33   **THE ACCURACY OF THE 3T3 AND NHK NRU TEST METHODS**

34

35   The aim of the procedures and analyses presented in this section is to identify the most

36   appropriate *in vivo* rodent toxicity data with which to compare the *in vitro* cytotoxicity data.

37   The *in vitro* NRU cytotoxicity test methods are intended to be used in a weight of evidence

38   approach to determine the starting dose for the *in vivo* acute oral systemic toxicity test

39   methods using rodents. Thus, rodent LD_{50} values from acute oral systemic toxicity tests are

40   the most appropriate reference data for the *in vitro* NRU IC_{50} values.   This section describes

41   the methods for identifying and evaluating the most appropriate rodent LD_{50} data to use for

42   determining reference LD_{50} values for the 72 reference substances tested in the

43   NICEATM/ECVAM validation study. These *in vivo* rodent toxicity reference values will be

44   used in **Section 6** to establish the accuracy of the *in vitro* IC_{50} data from the 3T3 and NHK

45   NRU test methods for predicting LD_{50} values from rodent acute oral systemic toxicity tests.

46

47   **4.1   Methods Used to Determine *In Vivo* Rodent Toxicity Reference Values**

48

49   4.1.1   Identification of Candidate *In Vivo* Rodent Toxicity Reference Data

50   No animal experiments were performed to obtain *in vivo* reference data for acute oral

51   systemic toxicity. To identify LD_{50} reference data for the 72 reference substances, rat oral

52   LD_{50} data were located through literature searches, secondary references, and electronic

53   database searches. PubMed and ISI Web of Science® searches were conducted using each

54   chemical name and "lethal dose 50." Secondary sources included NTP technical reports,

55   Toxicological Profiles from the Agency for Toxic Substances and Disease Registry

56   (ATSDR), Cosmetic Ingredient Reviews by the Cosmetics Industry Council, pesticide

57   handbooks, Merck Index, and various other summary sources. **Table 4-1** lists databases

58   searched via the Internet to locate references for rat oral LD_{50} values. Rat LD_{50} data were

59   preferred because the current oral acute toxicity test guidelines recommend using rats (OECD

60   2001a, c, d; EPA 2002a). Taking the same approach used for the Registry of Cytotoxicity

61   (RC), mouse LD_{50} data were sought for a particular chemical if rat LD_{50} values could not be

62   located. [The RC is a database of acute oral LD_{50} values for rats and mice, obtained from

63   RTECS$^®$ and IC$_{50}$ values from *in vitro* cytotoxicity assays using multiple cell lines and

64   cytotoxicity endpoints for chemicals with known molecular weights (Halle 1998).]

65

Table 4-1   Internet Accessible Databases with LD$_{50}$ Information

Database	Sponsor
Agency for Toxic Substances and Disease Registry (ATSDR)	U.S. Department of Health and Human Services (DHHS)
Center for Drug Evaluation and Research (CDER)	U.S. Food and Drug Administration (FDA)
CHEMFINDER	CambridgeSoft Corporation
Chemical Carcinogenesis Research Information System (CCRIS) National Cancer Institute (NCI) Website	NCI; National Institutes of Health (NIH); DHHS
Chemical Evaluation Search and Retrieval System (CESARS)	Michigan Department of Natural Resources; Ontario Ministry of the Environment; CCOHS CHEMpendium™
Chemical Hazard Response (CHRIS)	U.S. Coast Guard
Chemical Ingredients Database	U.S. Environmental Protection Agency (EPA) Office of Pesticide Programs (OPP); California EPA Department of Pesticide Regulation
CHEMINDEX CHEMINFO	Canadian Centre for Occupational Health and Safety (CCOHS) CHEMpendium™
ChemRTK High Production Volume (HPV) Challenge Program OPPT Chemical Fact Sheets Chemical Information Collection and Data Development	EPA Office of Pollution Prevention and Toxics (OPPT)
CIS Chemical Information	World Health Organization (WHO) International Programme on Chemical Safety (IPCS); CCOHS; International Labour Organisation (ILO) Occupational Safety and Health Information Centre (CIS)
Concise International Chemical Assessment Documents (CICADS)	WHO IPCS; CCOHS; ILO; United Nations Environment Programme (UNEP)
Consumer Product Safety Commission Website	U.S. Consumer Product Safety Commission (CPSC)
Deutsches Institut fur Medizinische Dokumentation und Information (DIMDI) [The German Institute for Medical Documentation and Information] Registry of Cytotoxicity (RC)	Zentralstelle zur Erfassung und Bewertungvon Ersatz- und Erganzungsmethoden zum Tierversuch (ZEBET) [German Centre for the Documentation and Validation of Alternative Methods]
Developmental and Reproductive Toxicology/Environmental Teratology Information Center (DART$^®$/ETIC)	EPA; The National Library of Medicine (NLM); The National Institute of Environmental Health Sciences (NIEHS); National Center for Toxicological Research (NCTR)
Emergency Response Guidebook (ERG 2000)	Transport Canada; U.S. Department of Transportation (DOT); Secretariat of Communications and Transportation of Mexico
Environmental Health Criteria (EHC) monographs Health and Safety Guides (HSG) International Agency for Research on Cancer (IARC)	WHO IPCS; CCOHS
European Centre for the Validation of Alternative Methods (ECVAM) Scientific Information Service (ECVAM SIS)	European Commission Joint Research Centre
HAZARDTEXT$^®$; MEDITEXT$^®$; INFOTEXT$^®$; SARATEXT$^®$; REPROTEXT$^®$; REPROTOX$^®$	TOMES Plus$^®$, MICROMEDEX, Greenwood Village, CO

Table 4-1   Internet Accessible Databases with LD$_{50}$ Information

Database	Sponsor
Integrated Risk Information System (IRIS)	EPA Office of Research and Development (ORD)
International Chemical Safety Cards (ICSC) IPCS/EC Evaluation of Antidotes Series	WHO IPCS; CCOHS; Commission of the European Union
International Uniform Chemical Information Database (IUCLID)	European Chemicals Bureau
Joint Expert Committee on Food Additives (JECFA) Joint Meeting on Pesticide Residues (JMPR) Pesticide Data Sheets (PDSs)	WHO IPCS; CCOHS; Food and Agriculture Organization (FAO) of the United Nations
Material Safety Data Sheets (MSDS)	Interactive Living Paradigms, Incorporated
Multicentre Evaluation of In Vitro Cytotoxicity (MEIC)	Scandinavian Society for Cell Toxicology
National Toxicology Program (NTP) Chemical Health and Safety Database	NIEHS
National Transportation Library	DOT
New Jersey Hazardous Substance Fact Sheets	New Jersey Department of Health and Senior Services
Oil and Hazardous Materials/Technical Assistance Data System (OHM/TADS)	EPA Office of Waste and Water Management
Organisation for Economic Co-operation and Development (OECD) Screening Information Data Sets (SIDS)	IPCS; CCOHS; International Register of Potentially Toxic Chemicals (IRPTC); UNEP
Pesticide Action Network Pesticide Database	Pesticide Action Network North America
Pesticide Product Information System (PPIS)	EPA Office of Pesticide Programs (OPP)
Poisons Information Monographs (PIMs)	IPCS; CCOHS
Registry of Toxic Effects of Chemical Substances (RTECS®) NIOSH Pocket Guide to Chemical Hazards	National Institute for Occupational Safety and Health (NIOSH)
SCORECARD	Environmental Defense
The EXtension TOXicology NETwork (EXTOXNET)	University of California, Davis; Oregon State University; Michigan State University; Cornell University; University of Idaho
The Right-to-Know Network (RTK NET)	Office of Management and Budget Watch; Center for Public Data access
Toxic Chemical Release Inventory (TRI) GENE-TOX	The National Library of Medicine (NLM)
Toxic Substances Control Act Test Submissions (TSCATS)	EPA OPPT
TOXLINE® Hazardous Substances Data Bank (HSDB) ChemIDplus	NLM (TOXNET)

66

67   A total of 195 LD$_{50}$ references retrieved through these searches were reviewed and evaluated.

68   Information regarding the materials and methods used to derive the 491 LD$_{50}$ values reported

69   by these references were compiled and are provided in **Appendix H-1** in a spreadsheet

70   format. **Appendix H-2** provides a narrative characterization and evaluation of the values.

71

72 4.1.2   Criteria Used to Select Candidate *In Vivo* Rodent Toxicity Data for Determination

73   of Reference Values

74 From the data retrieved, the goal was to derive a set of high quality reference LD_{50} values

75 (i.e., data that were collected using standardized protocols, accompanied by documentation

76 showing that established testing procedures were followed in compliance with national and

77 international GLP guidelines [OECD 1998; FDA 2003; EPA 2003a,b]). After a review of the

78 collected data, the SMT determined that a requirement for GLP compliance would eliminate

79 99% (452 of the 459 values remaining after exclusion of 30 duplicate values and two

80 erroneous values) of the oral LD_{50} values, since only seven had been obtained in compliance

81 with GLP guidelines. GLP-compliant studies were found for only four of the 72 (6%)

82 reference substances.

83

84 The SMT then considered limiting the selection of LD_{50} values to those from studies that

85 used the type of animals recommended by the current oral acute toxicity test guidelines, since

86 these guidelines will be used for future acute systemic toxicity testing. The current

87 guidelines recommend using young adult rats, 8-12 weeks of age, of a common laboratory

88 strain and the most sensitive sex (OECD 2001a, c, d; EPA 2002a). Female animals are

89 suggested if there is no information on which to determine the most sensitive sex. Selecting

90 LD_{50} values from animals that fit this description yielded a limited number of values. Only

91 3% (14/459) of the oral LD_{50} values were determined using 8-12 week old female laboratory

92 rats. Another 15 LD_{50} values were obtained with female rats in an appropriate weight range

93 (~ 176-250 g according to Charles River [http://www.criver.com], Harlan

94 [http://www.harlan.com/us/index.htm], and Taconic Farms

95 [http://www.taconic.com/anmodels/spragued.htm] websites) for that age. Thus, only 6%

96 (29/459) of the LD_{50} values in the database, covering 21 of the 72 reference substances (29

97 %), were obtained from studies that used the strain, sex, and age of rats recommended by

98 current test guidelines (OECD 2001a; EPA 2002a).

99

100 *Final Exclusion Criteria*

101 Since so few studies met the initial criteria (i.e., GLP compliance and use of animals

102 recommended by current acute oral toxicity test guidelines), the database was reviewed and

103 evaluated to derive alternative criteria for the development of reference LD_{50} values. For this

104 evaluation, the SMT looked for commonalities among the data records that, when selected,

105 provided a comparable data set for each chemical. Review of the available data indicated

106 that the majority of acute oral toxicity tests were conducted with unanesthetized young adult

107 laboratory rats of both genders, to which chemicals were administered by gavage. Thus, to

108 compile a homogenous set of reference LD_{50} values for each chemical, the selection process

109 was revised to exclude studies that reflected the following, less typical, materials and

110 methods:

111 • feral rats

112 • rats < 4 weeks of age

113 • anesthetized rats

114 • test chemical administered in food or capsule

115 • LD_{50} reported as a range or inequality

116

117 Data from feral rats were excluded, since the health status of these animals was uncertain.

118 All laboratory rat strains/stocks were deemed acceptable, since they were expected to be

119 healthy and provided with adequate care and housing during testing. Data from neonates or

120 weanlings were excluded since their sensitivity to chemical toxicity may differ from that of

121 adults. Four weeks was considered the minimum acceptable age, since rats are weaned at

122 about 3 weeks of age (Barrow 2000). Data from feeding experiments or experiments that

123 involved administration of the chemical in capsules were also excluded, since gavage is the

124 most common mode of administration for acute oral studies and the rate of gastrointestinal

125 absorption for these methods is likely to be different (Nebendahl 2000). Since LD_{50} point

126 estimates are required for the prediction model, LD_{50} values reported as ranges or inequalities

127 were considered unacceptable.

128

129 *Assumptions*

130 The level of detail for materials and methods for the LD_{50} studies varied greatly. Some

131 studies reported only the use of white rats while other acute oral toxicity studies provided

132 complete information on stock/strain, gender, and age of animals; the number of animals

133 tested per dosing group, method of administration, doses administered, clinical signs, and

134   times of death. To use as much of the available data as possible, the following assumptions

135   were made if a study report did not declare otherwise.

- The rats were assumed to have been young adults of a common laboratory
  strain.
- The rats were assumed to have been unanesthetized.
- The oral route of administration was by gavage.

141   *Calculation of Reference Values*

142   If there were multiple acceptable LD_{50} values for a chemical after the application of the

143   exclusionary criteria, outliers at the 99% level (Dixon and Massey 1981) were excluded. A

144   geometric mean and 95% confidence limits were calculated from the remaining values to

145   serve as the reference LD_{50}. A geometric mean is the antilog of the mean of the logarithm of

146   the values and it is less affected by extreme values than the arithmetic mean. Use of a

147   geometric mean corresponds with the approach used for the RC regression to obtain a single

148   IC_{50} value from multiple IC_{50} values (Halle 1998) and with the approach used to derive the

149   IC_{50} value for each chemical for the *in vitro - in vivo* regressions for the NICEATM/ECVAM

150   validation study (see **Section 6**).

152   In addition to the statistical evaluation of outliers, an extreme value, which was not a

153   statistical outlier, for trichloroacetic acid was also evaluated based on biological plausibility.

154   There were five LD_{50} values that ranged from 400-8900 mg/kg after applying the

155   exclusionary criteria for trichloroacetic acid. The lowest value of 400 mg/kg was rejected as

156   biologically implausible since up to 1000 mg/kg/day has been used in chronic rodent

157   carcinogenicity studies (EPA 1996).

159   *Use of Rat and Mouse Data*

160   If no rat oral LD_{50} values could be found for a chemical, mouse oral LD_{50} values were

161   located, retrieved, and evaluated by the same method as that used for rat oral values.

162   Although a model using entirely rat data or entirely mouse data would be preferable, the use

163   of mouse values was considered to be justified by a significant correlation of rat and mouse

164   oral LD_{50} values reported by Ekwall et al. (1998a) for the 50 chemicals tested in the MEIC

165   study. Using values from RTECS[®], Ekwall et al. (1998a) reported a coefficient of

166   determination, R^2, of 0.85 for a linear regression analysis of rat LD_{50} - mouse LD_{50}.

167   Furthermore, Halle (1998) compared IC_{50} - LD_{50} linear regressions with 285 rat values and

168   242 mice values and found no significant difference in the intercepts or slopes.

169

170   A correlation of the 173 chemicals in the RC that had both rat and mouse LD_{50} values is

171   shown in **Figure 4-1**. A Spearman correlation analysis of the log transformed rat and mouse

172   data yielded a significant correlation (p< 0.0001) with r_s = 0.88.

173

174   **Figure 4-1   Correlation of Rat and Mouse LD_{50} Values for 173 RC Chemicals**

175

176   The diagonal line shows the 1:1 relationship.

177

178

179

180
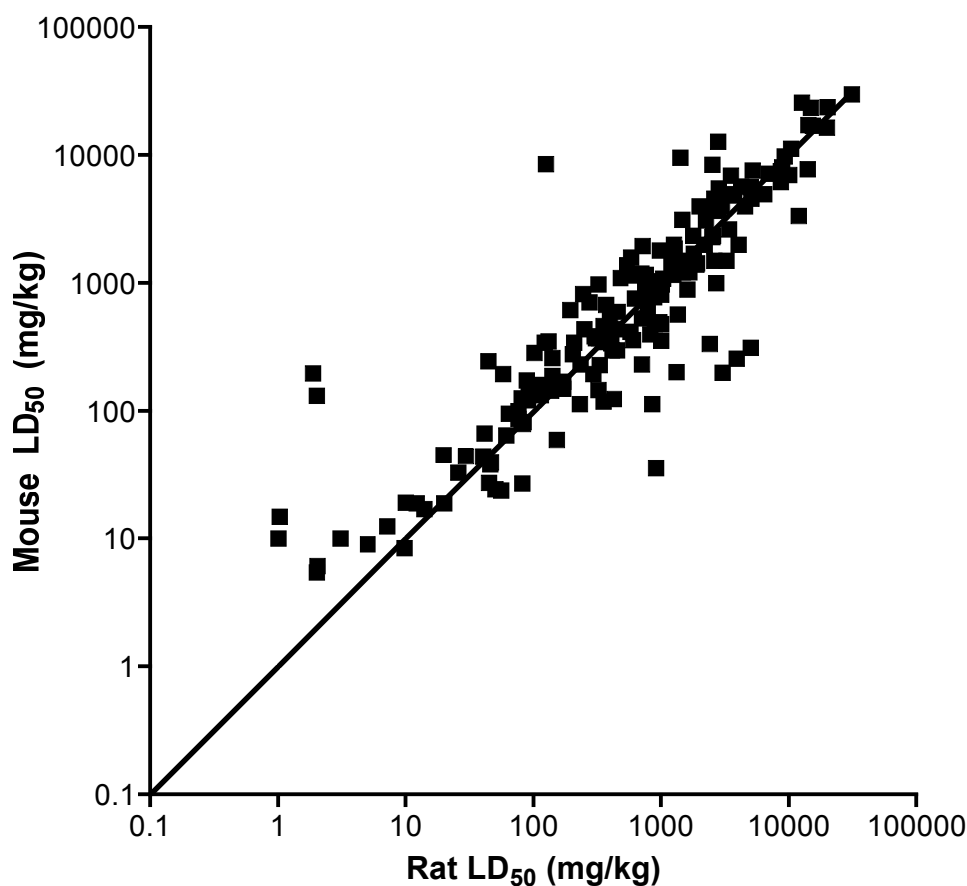
181   **4.2   Final *In Vivo* Rodent Toxicity Reference Values**

182

183   After the application of the exclusionary criteria, there were 385 acceptable LD_{50} values with

184   which to calculate reference values. **Table 4-2** shows the reference LD_{50} value for each

185   reference substance in ascending order. The reference values are the geometric means of the

186   acceptable LD_{50} values. Also shown for each substance are the 95% confidence limits

187   around the mean, the ratio of the maximum to the minimum acceptable value, the number of

188   LD_{50} values used to calculate the reference value, the number of LD_{50} values available (not

189   including duplicate values or the erroneous values for acetylsalicylic acid and sodium

190   oxalate), and the LD_{50} initially used for hazard category (often referred to as "toxicity" or

191   "LD_{50}" category) classification of the substance (see **Table 3-2**). Ratios for the maximum to

192   minimum LD_{50} values ranged from 1.0 to 25.9. The average ratio was 4.1. Six of the 62

193   reference substances for which ratios were calculated had ratios greater than one order of

194   magnitude: triethylenemelamine, parathion, busulfan, triphenyltin hydroxide, phenol, and

195   trichloroacetic acid. Three of these substances, triethylenemelamine, parathion, and

196   busulfan, were in the two highest toxicity categories (i.e., $LD_{50} \leq 50$ mg/kg).

197

198   **Table 4-2** shows the reference substances grouped by GHS acute oral toxicity category (UN

199   2005) using the reference LD_{50} values. The initial categorization for this study, which used

200   the LD_{50} values in the far right column of **Table 4-2** (i.e., values reported in **Table 3-2**,

201   which come from the RC unless otherwise specified), placed 12 substances in each toxicity

202   category. **Table 4-3** compares the number of substances in each GHS toxicity category

203   based on the reference LD_{50} values with the number of substances in each category based on

204   the initial LD_{50} values. **Table 4-3** shows that the initial and reference LD_{50} values placed

205   74% of the substances in the same GHS category. Compared with the initial LD_{50}, the

206   reference LD_{50} was higher for 25% of the substances and lower for 1% of the substances.

207

208

Table 4-2   Reference LD$_{50}$ Values by GHS Category[1]

GHS Category[1]/Chemical	Reference Oral LD$_{50}$[2] (mg/kg)	95% Confidence Interval[3] (mg/kg)	Reference Oral LD$_{50}$[2] (mmol/kg)	95% Confidence Interval[3] (mmol/kg)	Maximum: Minimum Value[3]	N Averaged[5]	Initial Rodent Oral LD$_{50}$[6] (mg/kg)
LD$_{50}$ ≤ 5 mg/kg (N = 7)							
Cycloheximide	2	NC	0.00711	NC	2.5	3	2
Phenylthiourea	3	NC	0.0197	NC	NC	1	3
Sodium selenate	3	NC	0.0159	NC	3.7	2	2[7]
Epinephrine bitartrate	4 (mouse)	NC	0.0196	NC	NC	1	4 (mouse)
Triethylenemelamine	4	1-25	0.0120	0.0037-0.12	13.0	4	1
Physostigmine	5	NC	0.0182	NC	NC	1	5[7]
Disulfoton	5	2-10	0.0182	0.009-0.036	5.5	6	2
5 < LD$_{50}$ ≤ 50 mg/kg (N = 12)							
Parathion	6	3-12	0.0209	0.010-0.041	16.7	10	2
Strychnine	6	NC	0.0188	NC	6.9	3	2[7]
Aminopterin	7	NC	0.016	NC	NC	1	3 (mouse)
Potassium cyanide	7	5-10	0.111	0.077-0.15	2.0	7	10
Busulfan	12	NC	0.049	0.008-0.38	15.3	4	2
Colchicine	15 (mouse)	NC	0.0375	NC	4.9	3	6 (mouse)
Thallium I sulfate	25	NC	0.0495	NC	NC	1	29 (mouse)
Arsenic III trioxide	25	10-64	0.127	0.050-0.32	6.3	5	20
Endosulfan	28	NC	0.068	NC	2.4	2	18[7]
Digoxin	28	NC	0.0362	NC	NC	1	18 (mouse)
Mercury II chloride	40	27-60	0.148	0.010-0.22	7.7	10	1
Sodium arsenite	44	36-53	0.336	0.28-0.40	1.5	5	41[7]
50 < LD$_{50}$ ≤ 300 mg/kg (N =12)							
Sodium dichromate dihydrate	51	44-58	0.193	0.17-0.22	1.9	11	50
Dichlorvos	59	40-88	0.266	0.18-0.40	5.7	9	17[7]
Nicotine	70	68-72	0.430	0.42-0.44	1.0	4	50
Fenpropathrin	76	57-100	0.217	0.16-0.29	3.4	9	18[7]
Hexachlorophene	82	68-98	0.202	0.17-0.24	3.8	19	61
Paraquat	93	65-132	0.498	0.35-0.71	2.0	5	58
Lindane	100	78-129	0.344	0.27-0.44	1.4	4	76
Verapamil HCl	111	NC	0.226	NC	1.1	2	108
Sodium I fluoride	127	92-175	3.020	2.19-4.16	4.4	12	180

Table 4-2　　Reference LD$_{50}$ Values by GHS Category[1]

GHS Category[1]/Chemical	Reference Oral LD$_{50}$[2] (mg/kg)	95% Confidence Interval[3] (mg/kg)	Reference Oral LD$_{50}$[2] (mmol/kg)	95% Confidence Interval[3] (mmol/kg)	Maximum: Minimum Value[3]	N Averaged[5]	Initial Rodent Oral LD$_{50}$[6] (mg/kg)
Cadmium II chloride	135	88-208	0.738	0.48-1.14	2.4	5	88
Diquat dibromide	160	NC	0.466	NC	1.9	3	231
Phenobarbital	224	NC	0.966	NC	2.0	3	163
300 < LD$_{50}$ ≤ 2000 mg/kg (N = 16)							
Caffeine	310	256-374	1.59	1.32-1.93	2.5	10	192
Triphenyltin hydroxide	329	208-520	0.896	0.57-1.42	25.9	15	44
Haloperidol	330	NC	0.877	NC	6.6	2	128[7]
Amitriptyline HCl	348	NC	1.18	NC	1.2	2	319
Propranolol HCl	466	NC	1.575	NC	NC	1	470 (mouse)
Cupric sulfate * 5 H2O	474	269-836	1.90	1.08-3.35	4.1	6	300
Phenol	548	434-692	5.82	4.82-7.68	4.7	14	414
Lithium carbonate	590	479-728	7.98	6.5-9.9	1.4	4	1187 (mouse; sulfate salt)
Glutethimide	600	NC	2.76	NC	NC	1	600
Sodium oxalate	633	NC	4.724	NC	1.3	2	155 (mouse)[8]
Chloral hydrate	638	391-1040	3.86	2.36-6.29	1.8	4	479
Atropine sulfate	819	641-1045	1.21	0.95-1.54	1.9	7	623
Valproic acid	995	NC	6.91	NC	2.2	2	670[7]
Meprobamate	1387	1291-1489	6.35	5.92-6.82	1.2	6	794[7]
Acetylsalicylic acid	1506	1224-1854	8.36	6.8-10.3	4.6	14	1000
Procainamide HCl	1950	NC	8.286	NC	NC	1	1950[7]
2000 < LD$_{50}$ ≤ 5000 mg/kg (N = 11)							
Acetaminophen	2163	NC	14.3	NC	1.2	2	2404
Potassium I chloride	2799	NC	37.6	NC	1.2	2	2602
Carbamazepine	2805	NC	11.9	NC	2.1	2	1957[7]
Boric aid	3426	2617-4486	55.4	42.3-72.6	1.9	6	2660[7]
5-Aminosalicylic acid	3429	NC	22.4	NC	1.5	2	7749 (mouse)
Chloramphenicol	3491	NC	10.8	NC	2.0	3	3393
Acetonitrile	3598	2951-4375	87.6	71.9-107	6.2	26	3798
Lactic acid	3639	NC	40.3	NC	1.1	2	3730
Carbon tetrachloride	3783	3024-4732	24.6	20-31	4.3	15	2799
Sodium chloride	4046	2917-5623	69.3	50-96	2.0	5	2998

Table 4-2   Reference LD$_{50}$ Values by GHS Category[1]

GHS Category[1]/Chemical	Reference Oral LD$_{50}$[2] (mg/kg)	95% Confidence Interval[3] (mg/kg)	Reference Oral LD$_{50}$[2] (mmol/kg)	95% Confidence Interval[3] (mmol/kg)	Maximum: Minimum Value[3]	N Averaged[5]	Initial Rodent Oral LD$_{50}$[6] (mg/kg)
Xylene	4667	1294-16827	43.9	12-158	5.6	4	4300
LD$_{50}$ > 5000 mg/kg (N = 14)							
2-Propanol	5105	4624-5636	84.9	77-94	1.3	6	5843
Trichloroacetic acid	5229	2745-9961	32.0	16.8-61.0	2.7	4	4999
Dimethylformamide	5309	3548-7925	72.6	49-108	2.6	6	2800
Citric Acid	5929	NC	30.9	NC	3.9	2	3000[7]
Gibberellic acid	6040	NC	17.4	NC	1.1	2	6305
Propylparaben	6332 (mouse)	NC	35.1	NC	NC	1	6326 (mouse)
Ethylene glycol	7161	6266-8204	115.4	101-132	2.5	16	8567
Methanol	8710	6223-12218	272	194-381	2.3	6	13012
Dibutylphthalate	8892	6180-12794	31.9	22-46	1.7	4	11998
Diethylphthalate	9311	NC	41.9	NC	1.2	2	8602
Sodium hypochlorite	10328	NC	62.8	NC	1.6	2	8910[9]
Ethanol	11324	8610-14894	245.7	187-323	2.5	8	14008
1,1,1-Trichloroethane	12078	10000-14588	90.5	75-109	1.7	6	10298
Glycerol	19770	10495-37154	215	114-403	2.2	4	12691

[1]GHS- Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005). Chemicals categorized using reference oral LD$_{50}$.
[2]Based on a geometric mean of acceptable LD$_{50}$ values from laboratory rats unless otherwise specified.
[3]For the geometric mean of the acceptable LD$_{50}$ values.
[4]Ratio of minimum acceptable LD$_{50}$ to maximum acceptable LD$_{50}$
[5]Number of values used for geometric mean.
[6]Values rounded to the nearest one; from the RC unless otherwise specified; rat data unless otherwise specified.
[7]RTECS® (MDL Information Systems 2002).
[8]RC reference for rat oral LD$_{50}$ of 155 mg/kg is Shrivastava et al. (1992), which references Klinger and Kersten (1961). Klinger and Kersten (1961) indicates the value was determined by intraperitoneal administration to mice.
[9]NLM (2002).
Abbreviations: NC – Not calculated. N was three or less and considered too small for a meaningful result.

221   The reference LD_{50} values caused the reclassification of 19 reference substances (i.e., the

222   sum of the numbers in the mismatching cells in **Table 4-3)**. Seven substances remain in the

223   lowest LD_{50} category (i.e., $LD_{50} \leq 5$ mg/kg). Five substances originally in this category

224   (aminopterin, mercury chloride, busulfan, parathion, and strychnine) moved to the next

225   higher category ($5 < LD_{50} \leq 50$ mg/kg) due the change in the reference LD_{50} values. In the 5

226    $< LD_{50} \leq 50$ mg/kg category, four substances (dichlorvos, fenpropathrin, sodium dichromate

227   dihydrate, and nicotine) moved to the next higher LD_{50} category ($50 < LD_{50} \leq 300$ mg/kg)

228   and one substance (triphenyltin hydroxide) moved up two categories to $300 < LD_{50} \leq 2000$

229   mg/kg. In the $50 < LD_{50} \leq 300$ category, four substances (haloperidol, caffeine, copper

230   sulfate pentahydrate, and sodium oxalate) moved up to the next toxicity category ($300 <$

231    $LD_{50} \leq 2000$ mg/kg). In the $300 < LD_{50} \leq 2000$ mg/kg category, only carbamazepine moved

232   up to the next toxicity category ($2000 < LD_{50} \leq 5000$ mg/kg). In the $2000 < LD_{50} \leq 5000$

233   mg/kg category, citric acid, trichloroacetic acid and dimethylformamide moved up to the next

234   higher LD_{50} category ($LD_{50} > 5000$ mg/kg). In the $LD_{50} > 5000$ mg/kg category, 5-

235   aminosalicylic acid moved down into the $2000 < LD_{50} \leq 5000$ mg/kg category. 5-

236   Aminosalicylic acid was the only substance that moved to a lower LD_{50} (i.e., more toxic)

237   category.

238   **Table 4-3   GHS[1] Toxicity Category Matches for the Initial and Reference LD$_{50}$ Values[2]**

Initial LD$_{50}$ (mg/kg)	Reference LD$_{50}$						Total	Category Match	Reference LD$_{50}$ Lower	Reference LD$_{50}$ Higher
	≤ 5	5-50	50 - 300	300 - 2000	2000 - 5000	> 5000				
≤ 5	7	5	0	0	0	0	12	58%	0%	42%
5-50	0	7	4	1	0	0	12	58%	0%	42%
50 - 300	0	0	8	4	0	0	12	67%	0%	33%
300 - 2000	0	0	0	11	1	0	12	92%	0%	8%
2000 - 5000	0	0	0	0	9	3	12	75%	0%	25%
> 5000	0	0	0	0	1	11	12	92%	8%	0%
Total	7	12	12	16	11	14	72	74%	1%	25%

239   [1]Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005):
240   ≤ 5: $LD_{50} \leq 5$ mg/kg
241   5 - 50: $5 < LD_{50} \leq 50$ mg/kg
242   50 - 300: $50 < LD_{50} \leq 300$ mg/kg
243   300 - 2000: $300 < LD_{50} \leq 2000$ mg/kg
244   2000 - 5000: $2000 < LD_{50} \leq 5000$ mg/kg
245   > 5000: $LD_{50} > 5000$ mg/kg
246   [2]Number of chemicals. Darkened cells show the number of chemicals for which the categories match.
247

248   **4.3   Relevant Toxicity Information for Humans**

249

250   The relevance of rodent acute systemic toxicity data to human lethality was assessed by the

251   MEIC program as a comparison to the evaluation of *in vitro* predictions of human acute

252   toxicity (Ekwall et al. 1998b). The MEIC program collected mouse and rat oral LD_{50} data

253   from RTECS[®] (Ekwall et al. 1998a). Mean lethal doses in humans were collected mainly

254   from handbooks containing human clinical toxicity information (Ekwall et al. 1998a). Data

255   from the handbooks were supplemented, when necessary, by an in-house compendium from

256   the Swedish Poisons Information Centre. Ekwall et al. (1998b) calculated least squares

257   linear regressions for the prediction of the mean human lethal doses by rat oral LD_{50} data and

258   by mouse oral LD_{50} data for the 50 MEIC chemicals using units of log mol/kg. Ekwall et al.

259   (1998b) reported $R^2 = 0.607$ for the rat LD_{50} prediction of mean human lethal doses and $R^2 =$

260   0.653 for the mouse LD_{50} prediction of mean human lethal doses.

261

262   The relevance of the NRU data collected in the NICEATM/ECVAM study to the prediction

263   of human acute toxicity will be addressed elsewhere by ECVAM.

264

265   **4.4   Accuracy and Reliability of the *In Vivo* Rodent Toxicity Reference Values**

266

267   Accuracy is the closeness of agreement between the results of an alternative test method and

268   an accepted reference test method (ICCVAM 2003). Since there is no accepted reference test

269   for the rodent acute oral toxicity test, the accuracy of the reference LD_{50} values for predicting

270   the oral LD_{50} in humans cannot be determined. Acute toxicity testing in rodents leads to a

271   relative ranking of the toxicity of chemicals for regulatory purposes. The reliability of the

272   reference LD_{50} values determined in this section may be judged by evaluating the range of

273   acceptable LD_{50} values for each chemical and by comparing the values (and their variability)

274   with other LD_{50} values.

275

276   *Variability Among the Acceptable LD_{50} Values*

277   The variability of the acceptable LD_{50} values used to calculate the reference value for each

278   reference substance was assessed by calculating the ratio of the maximum to the minimum

279   value (see **Table 4-2**). For the 62 reference substances with more than one acceptable LD_{50}

280   value, the average maximum:minimum ratio ranged from 1.1 to 25.9 with a mean of 4.3 and

281   a median of 2.2. The maximum:minimum ratios were greater than 10 for four substances:

282   triethylenemelamine, parathion, busulfan, and triphenyltin hydroxide.

283

284   The low LD_{50} values for triethylenemelamine, busulfan, and parathion may have contributed

285   to the high maximum:minimum ratios for these substances, since the range of values did not

286   seem to be extremely wide. The four LD_{50} values for triethylenemelamine ranged from 1 to

287   13 mg/kg, the four LD_{50} values for busulfan ranged from 1.9 to 29 mg/kg, and the 10 LD_{50}

288   values for parathion ranged from 1.8 to 30 mg/kg. **Table 4-4** shows the maximum:minimum

289   ratios by toxicity category. The substances in the higher toxicity categories (i.e., $LD_{50} \leq 50$

290   mg/kg) tended to have higher maximum:minimum LD_{50} ratios than substances in the lower

291   toxicity categories (i.e., $LD_{50} > 50$ mg/kg); however, there were also fewer substances in the

292   higher toxicity categories.

293

294   **Table 4-4   Maximum:Minimum LD_{50} Ratios by GHS[1] Toxicity Category**

GHS Category[1] (LD_{50} in mg/kg)	Mean Maximum:Minimum LD_{50} Ratio	Median Maximum:Minimum LD_{50} Ratio	Range of Maximum:Minimum LD_{50} Ratio	N
$LD_{50} \leq 5$	6.2	4.6	2.5 – 13.0	4
$5 < LD_{50} \leq 50$	7.1	6.3	2.0 - 16.7	9
$50 < LD_{50} \leq 300$	2.4	1.9	1.1 - 5.7	12
$300 < LD_{50} \leq 2000$	4.6	2.2	1.2 - 25.9	13
$2000 < LD_{50} \leq 5000$	2.6	2.0	1.2- 22.3	11
$LD_{50} > 5000$	2.3	2.3	1.1 - 3.9	13

295   [1]GHS-Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005).
296   N = number of chemicals with more than one acceptable LD_{50} value after application of the exclusion criteria in
297   **Section 4.1.2**.
298

299   *Comparison of Reference Values with RC Values*

300   The correspondence of the reference LD_{50} values with the LD_{50} values for the 58 validation

301   study reference substances in common with the RC are shown on a log scale in **Figure 4-2**.

302   A Spearman correlation analysis for the two sets of log transformed values yielded a

303   significant correlation ($p < 0.0001$) with a correlation coefficient, r_s, of 0.97. **Figure 4-2**

304   shows that the reference values tended to be higher than the RC LD_{50} values. The LD_{50}

305   values used in the RC were largely from the 1983/84 RTECS®, which publishes the lowest

306   LD_{50} value found for a particular chemical without regard to the source (i.e., from a primary

307   publication or a review) and without scientific review before publication. Thus, since the

308   reference LD_{50} values are based on the geometric mean from multiple studies, it is not

309   surprising that these values tended to be higher than those included in the RC database.

310

311   When comparing the reference LD_{50} values to the RC values, the substances with the largest

312   differences in LD_{50} were busulfan, triphenyltin hydroxide, and mercury chloride (see **Figure**

313   **4-2)**.

- The reference LD_{50} for busulfan was six times that of the RC value (12 mg/kg
  vs. 1.9 mg/kg). The RC value (i.e., the 1983/84 RTECS® value) was from a
  paper by Schmahl and Osswald (1970) in which they cited a rat oral LD_{50} of
  1.86 mg/kg. We also found rat oral LD_{50} values of 28 and 29 mg/kg for male
  and female Sprague-Dawley rats, respectively (Matsuno et al. 1971).
- The reference LD_{50} for triphenyltin hydroxide was 7.5 times the RC LD_{50} (329
  mg/kg vs. 44 mg/kg). The 15 LD_{50} values used to determine the reference value
  included the RC value and had a wide range, 44-1200 mg/kg. Due to the
  relatively large variation in the data, neither the highest nor the lowest values
  were statistical outliers.
- The reference LD_{50} for mercury chloride was 40 mg/kg, while the RC value was
  1 mg/kg. The RC value was from a summary document that reported the rat
  oral LD_{50} as a range of 1-5 mg/kg (Worthing and Walker 1991). Since it was
  reported as a range, it was excluded from the calculation of the reference value.
  The remaining 11 LD_{50} values ranged from 12 to 160 mg/kg. As previously
  stated, 160 mg/kg was an outlier compared to the other 10 values and therefore
  excluded from the calculation of the reference value.

331

332

333

334

335

336   **Figure 4-2   Correlation of LD$_{50}$ Values for the 58 RC Chemicals**
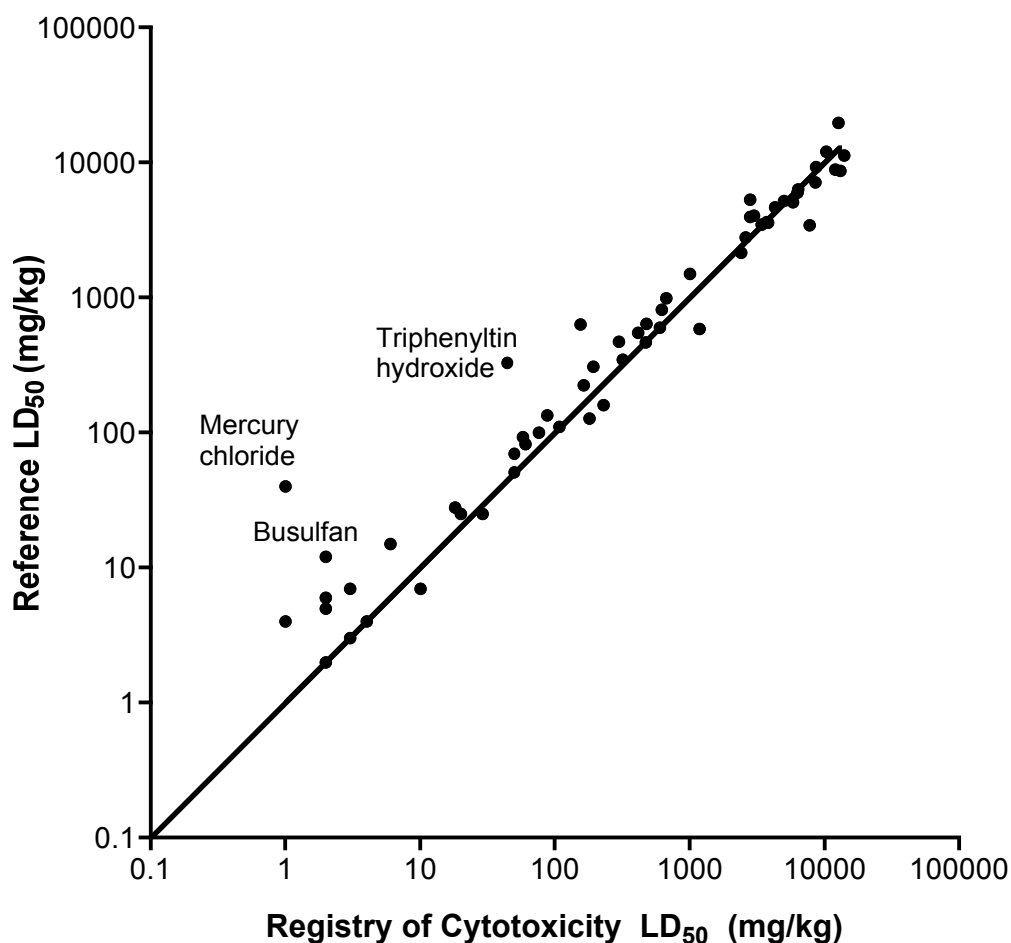
337
338   The diagonal line shows the 1:1 relationship.

339

340
341   *Comparison of the Variability Among Acceptable LD$_{50}$ Values to Other Studies*

342   When compared to other studies on the variation of acute oral LD$_{50}$ values, the variation

343   determined for 61 reference substances with multiple LD$_{50}$ values was not unusual. Weil and

344   Wright (1967) showed that even LD$_{50}$ values from multiple laboratories using exactly the

345   same protocol varied by as much as five-fold for the 10 substances they tested in eight

346   laboratories. In addition, they showed that allowing the laboratories to use their own

347   protocols for LD$_{50}$ determination produced data somewhat more variable, but the observed

348   differences were not reported. Another multicenter study that did not control the LD$_{50}$

349   protocols reported maximum:minimum ratios from 3.6 to 11.3 for five substances (Hunter et

350   al. 1979). The 65 participating laboratories in eight countries reported LD_{50} values ranging

351   from 44 to 5420 mg/kg for the five substances tested:

352   • Compound I/PCP   44 – 523 mg/kg

353   • Compound II/Sodium salicylate   800 - 4150 mg/kg

354   • Compound III/Aniline   350 – 1280 mg/kg

355   • Compound IV/Acetanilide   805 – 5420 mg/kg

356   • Compound V/Cadmium chloride   70 – 513 mg/kg

357

358   The results of a follow on study in which the same substances were tested by about 100

359   laboratories in 13 countries showed that adhering to a specific protocol reduced the range of

360   maximum:minimum LD_{50} ratios from 3.6 – 11.3 to 2.4 – 8.4 (Zbinden and Flury-Roversi

361   1981).

362

363   Although the LD_{50} data collected from the literature for the NICEATM/ECVAM validation

364   study used various strains, sexes, observation durations, and calculation methods for

365   estimating the LD_{50}, the variation in LD_{50} values for individual substances was similar to the

366   data by Hunter et al. (1979). The current study found six of the 61 substances with multiple

367   LD_{50} values had maximum:minimum LD_{50} values higher than that reported by Hunter et al.

368   (1979). Three of the reference substances: triethylenemelamine, parathion, and busulfan,

369   were in the lowest LD_{50} (i.e., highest toxicity categories). Hunter et al. (1979) also observed

370   that the largest variation was associated with the most toxic substances.

371

372   **4.5   Summary**

373

374   *In vivo* reference data for comparison with the *in vitro* NRU cytotoxicity data for the 72

375   substances were determined by analyzing rodent LD_{50} values identified by literature searches

376   and secondary references. Rat LD_{50} values were preferred, but when rat data could not be

377   located for three substances, mouse LD_{50} values were used. The 491 LD_{50} values located

378   consisted of 485 rat oral LD_{50} values and six mouse oral LD_{50} values. Identifying a high

379   quality data set determined under GLP guidelines was not possible since only 3% of the data

380   records were in compliance. Instead, a homogenous set of LD_{50} values for each substance

381   was identified by excluding studies that employed the following materials and methods:

382   • feral rats

383   • rats < 4 weeks of age

384   • anesthetized rats

385   • test chemical administered in food or capsule

386   • LD_{50} reported as a range or inequality

387

388   After analyzing the remaining acceptable data for outliers, the remaining 385 values were

389   used to determine *in vivo* reference values by calculating a geometric mean of the values for

390   each reference substance. The reference LD_{50} values for 20 substances varied enough from

391   the initial LD_{50} values, which came from the RC and other summary sources, that the

392   substances were classified into different GHS oral toxicity categories.

393

394   Since there is no reference test for the rodent oral LD_{50}, the accuracy of the reference values

395   for predicting the oral LD_{50} in humans could not be determined. The reliability of the

396   reference values was assessed by comparison to other evaluations of the performance of the

397   *in vivo* acute oral toxicity tests. Although the correlation of the reference values for the 58

398   RC chemicals with the RC LD_{50} was high ($r_s = 0.97$), the reference LD_{50} values tended to be

399   higher than the RC values. The maximum:minimum ratio of the acceptable values for the 62

400   reference substances that had more than one LD_{50} value ranged from 1.1 to 25.9. The

401   maximum:minimum ratios for four chemicals were greater than one order of magnitude.

402

402
403
404
405
406
407
408
409
410
411
412
413
414
415
416   ***[This Page Intentionally Left Blank]***
417

41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56   ***[This Page Intentionally Left Blank]***
57

58   **5.0   3T3 AND NHK NRU TEST METHOD DATA AND RESULTS**

59

60   This section presents *in vitro* IC_{50} data generated by testing coded reference substances using

61   the 3T3 and NHK NRU test method protocols. These IC_{50} values were used to evaluate the

62   accuracy (also known as concordance)(see **Section 6**) and reliability (interlaboratory

63   repeatability and reproducibility, intralaboratory reproducibility) (see **Section 7**) of these two

64   *in vitro* cytotoxicity test methods. **Section 5.1** summarizes protocol modifications and

65   revisions for each sequential phase of the validation study and examines whether such

66   changes affected the data. **Section 5.2** provides the data used for assessing the accuracy and

67   reliability of the 3T3 and NHK NRU protocols with a focus on PC data. **Section 5.3**

68   summarizes the statistical approaches used for data evaluation and **Section 5.4** provides

69   summaries of the acceptable 3T3 and NHK NRU test data for each reference substance

70   (average IC_{50} for each laboratory/test method). **Section 5.5** describes the "lot-to-lot"

71   consistency of the reference substances and adherence to GLP guidelines. **Section 5.6**

72   provides the study timeline, **Section 5.7** describes availability of test data, and **Section 5.8**

73   presents the solubility test data. The individual test data for both passing and failing tests

74   ($EXCEL^{®}$ and $PRISM^{®}$ files) and summary spreadsheets are available on compact disk(s).

75   Laboratory reports are also available on compact disk(s).

76

77   **5.1   3T3 and NHK NRU Test Method Protocols**

78

79   The protocols for the 3T3 and NHK NRU test methods used during Phase III laboratory

80   testing phase are a result of modifications and revisions of the *Guidance Document*

81   (ICCVAM 2001b) protocols and the optimization of the protocols used in the laboratory

82   evaluation phases (Phases Ia and Ib) and the laboratory qualification phase (Phase II).

83   **Figure 1-2** provides an outline of the study phases, as well as identifying where repeated

84   observations were carried out to permit protocol evaluation and comparison. The following

85   sections address the modifications of the protocols used in each phase and how those

86   modifications affected each data set (**Section 2** details the similarities and differences

87   between the two test method protocols).

88

89   5.1.1   Phase Ia: Laboratory Evaluation Phase

90   During Phase Ia, each testing laboratory established an historical database for the positive

91   control chemical, sodium lauryl sulfate (SLS). No reference substances were tested in this

92   phase. Ten concentration-response experiments were performed, with no more than two

93   experiments/day, and the resulting data were used to calculate the acceptable response limits

94   for use in Phase Ib testing.

95

96   **Section 2.6.1** summarizes issues that occurred during this phase and addresses protocol

97   changes made after the initiation of Phase Ia. The specific changes for both protocols are

98   summarized here along with the impact the change had on the test data. Changes made in the

99   protocols during Phase Ia were included in the Phase Ib protocols.

100

101   Protocol Changes and Impact on the Data

102   • *NR Dye Crystals:* Reduced the NR dye concentration for both cell types. No

103   subsequent tests failed due to NR crystal formation and no apparent impact on

104   the data was detected.

105   • *3T3 Cell Growth*: Modified cell culture conditions for 3T3 cells to improve cell

106   growth characteristics. No apparent impact on the data was detected.

107   • *NHK Cell Growth (96-well plates):* Removed the cell culture-refeeding step

108   performed prior to the reference substance application. SLS IC_{50} data were

109   similar whether the cells were refed or not refed. The change in the protocol did

110   not produce any observable impact on the data.

111   • *NHK Cell Growth (in culture flasks)*: FAL coated the culture flasks with

112   fibronectin-collagen prior to seeding thawed cells. No apparent impact on data

113   was detected.

114   • *OD Limits*: Eliminated the VC OD value range. The SMT accepted data from

115   tests that were out of the OD range if all other criteria were met. Test data were

116   not adversely affected by relaxing this criterion.

117   • *Dilution Factor*: The SMT accepted data generated using dilution factors other

118   than the recommended 1.47 for definitive tests if all other test acceptance

119   criteria were met. The use of smaller dilution factors generally increased the

120   number of points between 10 - 90% viability and the precision of the IC_{50}

121   calculation was improved.

122

123   5.1.2   Phase Ib: Laboratory Evaluation Phase

124   The purpose of Phase Ib was to determine whether the protocol revisions from Phase Ia were

125   effective in improving intra- and inter-laboratory reproducibility and to determine whether

126   the laboratories could obtain reproducible results when testing coded reference substances of

127   various toxicities. Three coded reference substances representing the full range of toxicity

128   were tested in Phase Ib: arsenic trioxide (high toxicity), propranolol (medium toxicity), and

129   ethylene glycol (low toxicity). Since Phase Ib was still part of the laboratory evaluation

130   phase, the SMT decided that testing just three substances was sufficient and the substances

131   did not need to represent all GHS toxicity categories. Each substance was tested at least once

132   in a range finding experiment and then in three acceptable definitive tests performed on three

133   different days.

134

135   **Section 2.6.2** summarizes the technical challenges that arose during this phase and addresses

136   protocol changes made after initiation of Phase Ib. This section (**5.1.2**) describes the specific

137   changes for the 3T3 and NHK NRU protocols along with the impact the changes had on the

138   test data.

139

140   *Protocol Changes and Impact on the Data*

141   • *NR Dye Crystals*: Reduced the concentration of NR in the 3T3 test method. The

142   OD values and SLS IC_{50} data were similar in four exploratory experiments

143   regardless of the NR concentration or the NRU incubation time tested. The

144   elimination of NRU crystals reduced the background OD values.

145   • *OD Range*: Used new OD ranges only for guidance (e.g., target values to assess

146   adequate cell growth) for the remainder of the study. This increased the number

147   of tests that met the acceptance criteria. Data were not adversely affected by the

148   removal of this criterion.

149   • *SLS IC_{50} Range*: Expanded the acceptance criterion range for the SLS IC_{50}.

150   This allowed additional positive control tests to meet the acceptance criteria and

151　　　　　　　thereby qualifying additional definitive tests as acceptable since they would

152　　　　　　　meet acceptance criteria and not fail simply because the PC failed.

153

154　5.1.3　　Phase II: Laboratory Qualification Phase

155　The results of Phase II determined whether the protocol revisions from Phase Ib were

156　effective in improving intra- and inter-laboratory reproducibility and whether the laboratories

157　could obtain reproducible results when testing a larger set of substances covering a wider

158　range of physical/chemical characteristics and toxicities than tested in Phase Ib. Nine coded

159　reference substances were analyzed: aminopterin, cadmium chloride, chloramphenicol,

160　colchicine, lithium carbonate, potassium chloride, 2-propanol, sodium fluoride, and sodium

161　selenate. These substances were common to the RC (with the exception of sodium selenate)

162　and were chosen because they fit the RC millimole regression line (i.e., were within the

163　acceptance intervals of the regression line). The RC is a database of acute oral LD_{50} values

164　for rats and mice obtained from RTECS$^{®}$ and IC_{50} values from *in vitro* cytotoxicity assays

165　using multiple cell lines and cytotoxicity endpoints for chemicals with known molecular

166　weights (Halle 1998). Sodium selenate, the non-RC chemical, was chosen because of its

167　high toxicity. Besides aminopterin, there were no other reference substances in the highest

168　toxicity category that were within the RC millimole regression acceptance intervals. Each

169　substance was tested at least once in a range finding experiment and then in three acceptable

170　definitive tests performed on different days during this phase.

171

172　**Sections 2.6.2** and **2.6.3** summarize the technical issues that arose during this phase and

173　address NRU protocol changes made prior to Phase II. This section (**5.1.3**) describes the

174　additional changes for both 3T3 and NHK NRU protocols along with the impact the changes

175　had on the test data.

176

177　*Protocol Changes and Impact on the Data*

178　　　　　　• *Blank Wells*: Added reference substance to blank wells of the test plate. There

179　　　　　　　was no apparent impact on test data.

180　　　　　　• *VC OD Range*: Eliminated the VC OD range as an acceptance criterion. There

181　　　　　　　was no apparent impact on test data.

182 • *Harmonization of Laboratory Techniques*: Made revisions to the Phase II
183 protocols as a result of the harmonization training by the testing laboratories
184 (see **Section 2.6.2**). There was no apparent impact on test data for IIVS and
185 ECBC but FAL data quality was improved.

186 • *3T3 Cell Seeding Density*: Added a range of cell seeding densities to be used by
187 the laboratories. No apparent impact on data was detected during this phase.

188 • *NHK Cell Growth from Cryopreservation*: Eliminated the use of fibronectin-
189 collagen coating and 80-cm^2 flasks for initial propagation of NHK cells. FAL
190 achieved better cell growth, obtained lower IC_{50} values for the PC, and achieved
191 better agreement of the mean SLS IC_{50} values compared to the other
192 laboratories.

193 • *Volatile Substances*: Added CO_2 permeable plate sealer use for control of
194 volatility in subsequent experiments (identified by cross contamination of the
195 control wells). The use of plate sealers for volatile substances was incorporated
196 into the Phase III protocols.

197 • *Hill Function*: Relaxed the Hill function criteria. Some tests that did not meet
198 the original criterion were accepted by the SMT after determining that even
199 though the curve fit was not optimum, the curve adequately conveyed the
200 toxicity of the substance.

201 • *Unusual Dose Response*: Revised the Hill function calculation to address
202 substances that produced a dose-response for which toxicity plateaued before
203 reaching 0% viability. This allowed for calculation of a more precise IC_{50} value
204 for such substances.

205 • *Positive Control IC_{50} Range*: Expanded the SLS IC_{50} acceptable range, which
206 resulted in additional tests in Phase II being acceptable. Expanding the PC
207 range reduced the number of retests of reference substances and thereby
208 qualifying additional definitive tests as acceptable since they would meet
209 acceptance criteria and not fail simply because the PC failed.

210

211

212

213   5.1.4   Phase III: Main Validation Phase

214   The purpose of Phase III was to generate high quality *in vitro* cytotoxicity data using the 3T3

215   and NHK NRU test methods with optimized test method protocols. Sixty coded reference

216   substances were tested (see **Table 5-3**); 46 of these were RC chemicals that covered a broad

217   range of toxicity. The substances in Phase III spanned all five GHS toxicity categories and

218   included unclassified substances. Each substance was tested at least once in a range finding

219   experiment and then in three acceptable definitive tests performed on different days. **Tables**

220   **5-3** and **5-4** provide summary data for the Phase III substances.

221

222   **Section 2.6.4** addresses protocol changes made before initiation of Phase III. This section

223   (**5.1.4**) describes the specific changes for both 3T3 and NHK NRU protocols along with the

224   impact the changes made on the test data.

225   - *Prequalification of NHK Culture Medium*: Included a protocol for prequalifying

226   NHK culture medium and supplements. This prevented the participating

227   laboratories from using medium and supplements that did not provide adequate

228   growth characteristics for NHK cells.

229   - *Stopping Rule for Testing*: Added this rule for chemicals that were insoluble

230   (i.e., solubility < 200 µg/mL) or could not achieve adequate toxicity over the

231   concentration range tested; this rule allowed testing to end for chemicals that

232   produced no IC_{50} data within three definitive tests. Chemicals that could not be

233   adequately tested by one or more laboratories are presented in **Table 5-1**. In all

234   three laboratories, carbon tetrachloride could not be adequately tested in either

235   3T3 or NHK cells while methanol could not be adequately tested in 3T3 cells.

236   - *Acceptable Range for Dose-Response Data Points:* Modified the test acceptance

237   criterion for the number of data points required on the toxicity curve. Changed

238   from requiring a minimum of two points (at least one point > 0% and ≤ 50%

239   viability and at least one point > 50% and < 100% viability) to one point > 0%

240   and < 100% viability if the smallest practical dilution factor was used (i.e., 1.21)

241   and all other test acceptance criteria were met. This reduced the number of

242   failed experiments without reducing the quality of the IC_{50} data.

243   • *R² Acceptance Criteria*: Rescinded the R^2 criterion for the fit of the Hill
244   function. The SMT determined that the R^2 criterion was best used to
245   characterize the reference chemical response curve shape rather than to establish
246   a criterion for test acceptability. This reduced the number of failed experiments
247   without reducing the quality of the IC_{50} data.
248   • *PC Acceptance Criteria*: Modified the PC acceptance criterion for Hill function
249   fit.
250   • *Hill Function Analysis*: Altered the PRISM® template for the Hill function
251   analysis to perform calculations for IC_x values in two ways: (1) constraining
252   Bottom parameter to zero and (2) fitting the Bottom parameter. As a result of
253   the changes and efforts by the laboratories to use dilution schemes that captured
254   the entire dose-response, very few tests in Phase III had $R^2 < 0.9$.
255   • *Biphasic Dose Response*: This aspect was added to the Phase III protocol so that
256   the Study Directors could make a decision about analyzing data from reference
257   substances with biphasic dose-responses (See **Section 2.6.3**).
258

259   **Table 5-1   Reference Substances Affected by Stopping Rule**
260

Reference Substance[1]	Testing Stopped -- No Data					
	3T3 NRU Test Method			NHK NRU Test Method		
	ECBC	FAL	IIVS	ECBC	FAL	IIVS
Carbon tetrachloride	X	X	X	X	X	X
Disulfoton		X				
Gibberellic acid		X				
Methanol	X	X	X	X		
1,1,1-Trichloroethane	X				X	X
Valproic acid			X			
Xylene	X	X		X	X	

261   [1]Substances that did not provide adequate cytotoxicity
262   ECBC: Edgewood Chemical Biological Center
263   FAL: FRAME Alternatives Laboratory
264   IIVS: Institute for In Vitro Sciences
265

266   **5.2   Data Obtained to Evaluate Accuracy and Reliability**
267

268   This section first presents the acceptable PC data from each laboratory for each phase of the
269   validation study and then presents the reference substance data for each phase. All test data,
270   both acceptable and unacceptable, are available on compact disk upon request. Accuracy

271   (concordance) and reliability assessments are provided in **Section 6** and **Section 7**,

272   respectively.

273

274   5.2.1   PC Data

275   A summary of the acceptable SLS IC_{50} data used to calculate quality control acceptance

276   limits for each experiment, by laboratory, to use in subsequent study phases, are shown in

277   **Table 5-2.**

278

279   *Phase Ib Acceptance Limits*

280   The acceptance limits for the SLS IC_{50} for Phase Ib testing were calculated using the Phase Ia

281   data. The data sets from each laboratory were examined for outliers using the method of

282   Massey and Dixon (1981), but none were identified. The acceptance limits for the SLS IC_{50}

283   values for each laboratory and test method were mean $\pm$ 2 SD since the SD is more

284   commonly used as a range than the 95% confidence limits.

285

286   *Phase II Acceptance Limits*

287   The IC_{50} values from the SLS tests from Phases Ia and Ib were used to calculate laboratory-

288   specific and test method-specific quality control acceptance limits for Phase II. Phase Ib

289   tests with SLS IC_{50} values outside of the acceptance limits were considered acceptable if they

290   met all other test acceptance criteria. For any day during which there was more than one SLS

291   test (for each test method and laboratory), the IC_{50} values were averaged to better reflect day-

292   to-day variation and avoid overweighting the overall mean with values from an individual

293   day. Extreme values were tested and removed if they were outliers at the 99% level and the

294   remaining values were used to calculate the mean $\pm$ 2.5 SD as the acceptance limits. The

295   acceptance limits were expanded from 2 SD in Phase Ib to 2.5 SD for Phase II to allow for

296   the fact that the limits tend to get narrower as more data are collected.

297   **Table 5-2   Positive Control (SLS) Data by Phase**

Study Phase	ECBC				FAL				IIVS			
	Mean IC50 (µg/mL)	Standard Deviation (µg/mL)	Acceptance Limits	N	Mean IC50 (µg/mL)	Standard Deviation (µg/mL)	Acceptance Limits	N	Mean IC50 (µg/mL)	Standard Deviation (µg/mL)	Acceptance Limits	N
3T3												
Ia[1]	38.3	4.71	28.8 – 47.7	15	42.3	8.56	25.2 – 59.5	25	40.9	3.19	34.5 – 47.3	12
Ib[2]	41.3	5.99	26.4 – 56.3	12	43.2	4.68	31.5 – 54.9	17	42.1	3.40	33.6 – 50.6	13
II[3]	41.2	4.20	30.8 – 51.6	29	45.9	7.50	27.2 – 64.7	36	40.6	3.50	31.8 – 49.3	21
III[4]	41.6	3.41	NA	65	41.1	6.23	NA	26	41.5	3.74	NA	22
NHK												
Ia[1]	4.03	1.32	1.40 – 6.67	15	7.45	3.07	1.34 – 13.6	18	3.68	0.555	2.57 – 4.79	30
Ib[2]	3.65	0.98	1.22 – 6.10	11	5.35	2.32	0[a] – 11.1	15	3.57	0.59	2.10 – 5.04	17
II[3]	3.59	1.41	0.07 – 7.11	22	3.20	1.05	0.57 – 5.82	15	3.78	0.73	1.94 – 5.61	26
III[4]	3.03	0.75	NA	57	3.45	0.90	NA	35	3.12	0.53	NA	20

298   [1]Values generated from Phase Ia data for PC acceptance criterion for Phase Ib; Acceptance limits = Mean ± 2 X standard deviation
299   [2]Values generated from Phases Ia and Ib data for PC acceptance criterion for Phase II; Acceptance limits = Mean ± 2.5 X standard deviation
300   [3]Values generated from Phases Ia, Ib, and II data for PC acceptance criterion for Phase III; Acceptance limits = Mean ± 2.5 X standard deviation
301   [4]Values generated from Phase III data.
302   [a]Calculation of lower limits actually yielded negative concentrations, so lower limit was placed at 0 and later revised to 0.1 µg/mL
303   NA = not applicable
304   ECBC: Edgewood Chemical Biological Center
305   FAL: FRAME Alternatives Laboratory
306   IIVS: Institute for In Vitro Sciences

307

308   *Phase III Acceptance Limits*

309   The IC_{50} values from the SLS tests from Phases I and II were used to calculate laboratory-

310   specific and test method-specific quality control acceptance limits for Phase III. The SLS

311   IC_{50} values outside of the acceptance limits were considered acceptable if the tests met all

312   other test acceptance criteria. For any day for which there was more than one SLS test (for

313   each test method and laboratory), the IC_{50} values were averaged to better reflect day-to-day

314   variation and avoid overweighting the overall mean with values from an individual day.

315   ANOVA was used to compare the Phase Ia, Ib and II data within each laboratory. For phases

316   that were not significantly different at $p < 0.05$, the IC_{50} data were used to calculate the mean

317   ± 2.5 SD as the acceptance limits for Phase III. The only laboratory/test method that showed

318   a significant difference between the phases was FAL using the NHK NRU test method ($p <$

319   0.0002). The difference was attributed to the changes in cell culture practices between

320   Phases Ib and II (see **Section 5.1.3**). Thus, for the NHK data at FAL, only the Phase II SLS

321   IC_{50} values were used to calculate the acceptance limits for Phase III.

322

323   The IC_{50} values from the SLS tests from Phase III are also presented in **Table 5-2**.

324

325   5.2.2   <u>Reference Substance Data</u>

326   All reference substance data from all laboratories are presented in **Appendix I**. **Tables 5-3**,

327   **5-4**, and **5-5** and **Figures 5-1 a-f** (3T3) and **5-2 a-f** (NHK) provide summary data for all

328   phases of the NICEATM/ECVAM validation study (see **Section 5.4**).

329

330   **5.3**   **Statistical Approaches to the Evaluation of 3T3 and NHK NRU Data**

331

332   Statistical approaches to data evaluation are reviewed in the following sections for each

333   phase of the NICEATM/ECVAM validation study. **Section 2.2.3** discusses the endpoint

334   measurements for the 3T3 and NHK NRU test methods. The mean OD values of the six

335   replicate values (six wells [minimum of four] in the 96-well plate) per test concentration

336   (eight concentrations/reference substance or PC) are used to determine relative cell viability

337   by calculating the specific concentration's percentage of the mean NRU of all VC values on

338   the same plate. The mean cell viability values generated from replicate wells for each

339   concentration are used to plot a toxicity curve (percent viability versus concentration) and the

340   IC_{50} value is determined from that curve.

341

342   5.3.1   <u>Statistical Analyses for Phase Ia</u>

343   The laboratories reported the IC_{50} results for SLS in μg/mL. The SMT used the results from

344   the acceptable tests to calculate means and SDs for each test method at each laboratory.

345

346   *Outlier Determination for Replicate Well Concentration Data*

347   During a review of the six replicate well OD data for the same concentration of a reference

348   substance, it was noted that extreme OD values sometimes occurred and that removal of

349   these "outlier values" frequently improved the fit of the Hill function for the concentration

350   cytotoxicity response curve. Concern was expressed that the outliers, if not excluded, may

351   create so much noise that the true cytotoxicity response might be obscured although there

352   was no discernable experimental reason for the outliers. Although it was recognized that

353   removal of extreme values reduced reported variability and might have altered the mean

354   value, an outlier test from Dixon and Massey (1981) was used to evaluate the consistency of

355   replicate well data. The SMT manually applied the outlier test to the Phase Ia data when

356   apparent extreme values were noted. If the extreme value was an outlier at the 99% level, it

357   was excluded from the data set, and the IC_{50} was recalculated. All data are available in the

358   data files provided by the laboratories, including the OD values in the excluded outlier value

359   wells. The protocol acceptance requirement of a minimum of four test wells per reference

360   substance concentration remained in effect.

361

362   *Curve Fit Criterion*

363   Upon visual review of the fit of the OD data to the Hill function curve, a curve fit criterion

364   was implemented as a test acceptance criterion. The SMT considered the fit of the

365   concentration-response curve to the Hill function to be acceptable when $R^2 > 0.9$. If $R^2 <$

366   0.8, then the fit was unacceptable and the data for that test was rejected. Curves with a fit of

367   $0.8 < R^2 < 0.9$ were evaluated visually (for goodness of fit) and accepted if the SMT

368   concluded that there were sufficient data points between 0 and 100% cytotoxicity and a

369   reasonable shape to the curve to calculate a reasonably accurate IC_{50}. Each test with a curve

370   fit in this range was analyzed individually (i.e., on a case-by-case basis) and no standard

371   criterion was developed to pass/fail such results. [Note: The use of R^2 was reevaluated in

372   Phases Ib and II and was eliminated as a test acceptance criterion for Phase III reference

373   substances. An R^2 value ≥ 0.85 was maintained as a test acceptance criterion for the PC.]

374   The R^2 criterion was implemented approximately two months after the laboratories

375   completed Phase Ia testing.

376

377   *Reproducibility Analyses for PC IC$_{50}$ Values*

378   To evaluate reproducibility of the IC$_{50}$ values for the PC for each test method, within and

379   between the laboratories, the SMT considered using the American Society of Testing and

380   Materials (ASTM) Standard E691-99, Standard Practice for Conducting an Interlaboratory

381   Study to Determine the Precision of a Test Method (ASTM 1999). This method uses two

382   statistics, h and k, to judge the consistency of means and variances between laboratories.

383   Since a minimum of six laboratories is required for this type of analysis, the SMT decided

384   that it could not be appropriately applied to three laboratories.

385

386   Therefore, the variability of the IC$_{50}$ data obtained for each test method and laboratory for the

387   PC was assessed using CV analysis and one-way analysis of variance (ANOVA). The CV

388   was calculated by dividing the SD by the arithmetic mean IC$_{50}$ value and then multiplying by

389   100. CV values were calculated for the acceptable tests within each laboratory. To compare

390   the variation among laboratories, CV was calculated from the mean IC$_{50}$ values from each

391   laboratory. Although no criterion for acceptable CV was determined for this study, ECVAM

392   has recently used CV < 30% as an acceptable CV range for both intra- and inter-laboratory

393   reproducibility (Zuang et al. 2002; Fentem et al. 2001). ECVAM usually applies the

394   criterion to the mean CV for all substances tested during the same phase. Although this CV

395   range is intended to reflect an acceptable maximum for normal biological variability, the

396   range is not supported by data.

397

398   For the ANOVA, IC$_{50}$ values were first converted to mM units and then log-transformed to

399   obtain normal distributions. One-way ANOVA was performed with SAS PROC GLM (SAS

400   Institute 1999; see **Appendix R1** for example SAS code). To be conservative with respect to

401   identifying laboratory differences, a significance level of $p < 0.01$ was used to test results

402   between the laboratories.

403

404   5.3.2   <u>Statistical Analyses for Phase Ib</u>

405   *Outlier Determination for Replicate Well Concentration Data*

406   For consistency of replicate well concentration data, the SMT applied the same outlier test

407   used for the Phase Ia data (Dixon and Massey 1981) when extreme OD values were noted. If

408   the extreme value was an outlier at the 99% level, it was excluded from the data set, and the

409   IC_{50} was recalculated. All data are available in the data files provided by the laboratories,

410   including the OD values in the excluded outlier value wells.

411

412

413   *Reproducibility Analyses for the Reference Substance IC_{50} Values*

414   A one-way ANOVA and CV analyses were used to assess test method reproducibility within

415   and across laboratories were performed as described in **Section 5.3.1**. When the ANOVA

416   detected significant differences among the laboratories ($p < 0.01$), contrast analyses were

417   performed to determine which laboratory was different from the others. The contrasts

418   compared the results of each laboratory with those of the other two laboratories. A

419   significant difference among the laboratories was indicated by $p < 0.01$.

420

421   5.3.3   <u>Statistical Analyses for Phase II</u>

422   *Outlier Determination for Replicate Well Concentration Data*

423   For consistency of replicate well concentration data, the outlier test from Dixon and Massey

424   (1981) was incorporated into the EXCEL® templates used by the laboratories to collect and

425   report data. Extreme values that were outliers at the 99% level were highlighted and the

426   Study Director was offered the option of removing the value from subsequent calculations

427   (for mean OD of the six replicates, % viability, IC_{50}, etc.).

428

429   *Reproducibility Analyses for Reference Substance IC_{50} Values*

430   CV values from the acceptable tests were used to calculate mean, SD, and CV for each

431   substance/test method/laboratory as described in **Section 5.3.2**. Intra- and inter-laboratory

432   reproducibility of IC$_{50}$ data, by test method, for the reference substances tested in Phases II

433   was also assessed using one-way ANOVA as described in **Section 5.3.2**.

434

435   *Comparison of 3T3 and NHK NRU Test Results to the RC Millimole Regression*

436   To compare the 3T3 and NHK NRU test results for the reference substances to those of the

437   RC millimole regression, the IC$_{50}$ values reported by the laboratories were transformed to

438   mM units for the calculation of geometric mean IC$_{50}$ values for each substance/test

439   method/laboratory. The log geometric mean IC$_{50}$ values were used with the RC LD$_{50}$ values

440   (see **Table 3-2**), after transformation to log mmol/kg units (see **Appendices J1** and **J3**), to

441   calculate least squares linear regressions for each test method and laboratory. Each of these

442   regressions was compared to the RC millimole regression using an F test with SAS PROC

443   REG (SAS Institute 1999; see **Appendix R2** for example SAS code). An F test with a

444   significance level of $p < 0.01$ was used to determine whether the joint comparison of slope

445   and intercept indicated that the laboratory regressions were significantly different from the

446   RC millimole regression.

447

448   5.3.4   <u>Statistical Analyses for Phase III</u>

449   *Outlier Determination for Replicate Well Concentration Data*

450   The laboratories used the outlier test at the 99% level (Dixon and Massey 1981) incorporated

451   into the EXCEL® templates to test for outlier values among replicate well concentration data.

452   The Study Director had the option of excluding the outliers from the data set, which were

453   highlighted by the template, from subsequent calculations. All data are available in the data

454   files provided by the laboratories, including the OD values in the excluded outlier value

455   wells.

456

457   *Reproducibility Analyses for the PC Data*

458   A number of analyses were performed to determine whether the SLS IC$_{50}$ values were

459   reproducible over the duration of the study (i.e., across study phases). The SLS IC$_{50}$ values

460   used to access variability were somewhat different from those shown in **Table 5-2**. To get an

461   assessment of the true variation of SLS IC$_{50}$ values, the reproducibility analyses included

462   IC$_{50}$ values from SLS tests that failed the test acceptance criterion for the IC$_{50}$ acceptance

463   limits in **Table 5-2** that were determined for each laboratory and study phase. These SLS

464   tests, however, passed all other test acceptance criteria. If more than one SLS test was

465   performed in a single day (for each test method and laboratory), the IC_{50} values were

466   averaged to determine a single IC_{50} for the day so that multiple results from a single day

467   would not overly influence the average for each phase. CV analyses were performed as

468   described in **Section 5.3.1** using the arithmetic mean IC_{50} values for each test method,

469   laboratory, and study phase.

470

471   For the remaining analyses of reproducibility, the IC_{50} values were first log-transformed to

472   obtain normal distributions. One-way ANOVAs were performed with SAS PROC GLM

473   (SAS Institute 1999; see **Appendix R1** for example SAS code) for each test method using

474   study phase and laboratory individually as explanatory variables. A significance level of p <

475   0.01 was used to test for a statistical difference among the laboratory and/or phase results.

476   To determine whether there was a linear time trend for the SLS IC_{50} data, linear regression

477   analyses using a least squares method were performed for each laboratory and test method

478   using SAS PROC REG (SAS Institute 1999). Time was expressed as an index for each test.

479   The index number of each test reflected its order of testing without respect to the time lapsing

480   between tests. The slopes of the linear regressions were statistically significant if p < 0.05.

481

482   *Reproducibility Analyses for the Reference Substance Data*

483   CV and one-way ANOVA analyses were performed to assess the intra- and inter-laboratory

484   reproducibility of the Phase III reference substance data as described in **Section 5.3.2**.

485

486   The geometric mean IC_{50} values were used to calculate least squares linear regression models

487   after log transforming the data. Linear regressions were fit for each test method and

488   laboratory using the log transformed reference LD_{50} values from **Table 4-2** in mmol/kg with

489   log IC_{50} in mM. To detect differences between the laboratory regressions, two models were

490   fit for each test method. The first model was a full model that included effects for laboratory

491   and interactions. This model generated a regression line for each laboratory. The second

492   model, the reduced model, assumed that one model fit all the laboratories. A goodness of fit

493   F test was performed to compare the full and reduced models for the two regressions for each

494   test method. A significance level of $p < 0.05$ was used to test whether the laboratory

495   regressions were significantly different from one another.

496

497   *Comparison of 3T3 and NHK NRU Test Results to the RC Regression*

498   The laboratory regressions for each test method were combined using the log geometric

499   mean of the geometric mean IC_{50} values from each laboratory and the reference log

500   transformed LD_{50} in mmol/kg. Another linear regression was calculated using the log

501   transformed IC_{50} and LD_{50} data from the RC for the 58 RC chemicals tested in the

502   NICEATM/ECVAM validation study. The regression for the 58 RC chemicals was

503   compared to the combined laboratory regressions for each test method using an F test to

504   compare slope and intercept (simultaneously). A $p < 0.01$ was used to indicated whether the

505   test method regressions were statistically different from the 58 chemical RC regression.

506

507   To assess accuracy of the regression models and the NRU test methods, the LD_{50} predictions

508   of the RC millimole regression and two additional regressions developed in **Section 6.2** were

509   used to assign predicted GHS acute oral toxicity category categories (see **Section 6.3**).

510   Accuracy was determined by calculating the proportion of chemicals for which the predicted

511   GHS toxicity category matched the *in vivo* GHS toxicity category. The LD_{50} predictions

512   from these regression models were also used to determine starting doses for acute systemic

513   toxicity test method simulations for the purpose calculating animal use and animal savings

514   using the NRU test methods. The simulation modeling methods and results for the UDP and

515   ATC methods are described in **Section 10**.

516

517   **5.4   Summary of Results**

518

519   **Table 5-3** the reference substance name, chemical class (classification based on the National

520   Library of Medicine's Medical Subject Heading [MeSH]), summary IC_{50} data (arithmetic

521   mean), standard deviations, and the number (N) of tests used to produce the values in the

522   study for both *in vitro* NRU cytotoxicity test methods. Data are categorized alphabetically

523   and by phase. The reference substance data are also shown on bar graphs in **Figures 5-1 a-f**

524   (3T3) and **5-2 a-f** (NHK) and the reference substances are ranked by IC_{50} values (lowest

525　value [most toxic] to highest value [least toxic]). The substances are divided into subgroups

526　for ease of fit to the graph size. **Appendices I-1** through **I-4** provide all test data (IC$_{50}$

527　values) from all laboratories for each cell type. **Tables 5-4** and **5-5** provide the geometric

528　IC$_{50}$ mean values for 3T3 and NHK (laboratories combined) and show the differences in the

529　values in orders of magnitude. The correlation of the mean IC$_{50}$ values for the 58 study

530　reference substances common to the RC database vs the RC IC$_{50}$ values is shown in **Figure**

531　**5-3** (3T3 NRU values) and **Figure 5-4** (NHK NRU values). **Table 5-7** contains summary

532　data for the solubility studies performed by the laboratories. **Table 5-8** lists the reference

533　substances that exhibited precipitate and/or volatility problems. **Appendix F** provides

534　physical, chemical, and biological information for all 72 reference substances.

535

Table 5-3   3T3 and NHK NRU Test Method Summary IC$_{50}$ Data from the Laboratories

Substance	Chemical Class[4]	3T3 NRU Test Method									NHK NRU Test Method									
		ECBC			FAL			IIVS			ECBC			FAL			IIVS			
		IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N	
Phase Ia																				
Sodium lauryl sulfate (SLS)	Alcohol	38.6	3.8	12	44.8	4.7	21	40.9	3.2	12	4.11	1.4	13	6.64	2.1	14	3.63	0.5	29	
Phase Ib																				
Arsenic III Trioxide	Arsenical	2.41	0.782	4	1.04	0.070	4	4.09	2.23	3	7.77	2.54	4	2.55	1.92	6	20.9	6.40	3	
Ethylene glycol	Alcohol	18325	1658	4	31650	7453	4	25900	3081	3	38000	4681	3	49800	4371	3	40000	5341	4	
Propranolol HCl	Alcohol	13.6	4.37	4	13.5	6.85	4	17.6	3.78	3	38.3	4.54	3	43.8	2.52	3	28.6	3.28	4	
Phase II																				
Aminopterin	Heterocyclic	0.005	0.001	3	0.012	0.005	3	0.005	0.001	3	889	182	3	545	42.2	3	611	70.7	2	
Cadmium II chloride	Cadmium compound	0.480	0.066	3	0.400	0.129	3	0.817	0.427	3	2.20	0.823	5	1.88	1.22	3	1.86	0.151	3	
Chloramphenicol	Alcohol	55.3	12.4	4	273	82.2	4	156	27.9	3	318	142	3	414	182	4	367	79.7	3	
Colchicine	Heterocyclic	0.021	0.002	4	0.093	0.042	3	0.028	0.0003	3	0.005	0.002	3	0.008	0.001	3	0.008	0.002	3	
Lithium I carbonate	Lithium compound	564	67.6	3	NA	NA	NA	NA	NA	NA	411	119	3	486	95.7	3	535	31.6	3	
Potassium I chloride	Potassium, chlorine compound	3352	468	4	3842	1198	5	3710	417	3	2560	432	3	2287	631	3	1990	161	3	
2-Propanol (Isopropyl alcohol)	Alcohol	2610	240	2	3970	139	3	4110	161	3	5263	583	3	4273	1139	3	7087	480	3	
Sodium I fluoride	Sodium, fluorine compound	61.3	5.55	3	96.1	17.7	3	82.0	5.81	3	48.7	6.92	3	39.7	9.61	3	53.7	6.82	4	

Table 5-3   **3T3 and NHK NRU Test Method Summary IC_{50} Data from the Laboratories**

Substance	Chemical Class[4]	3T3 NRU Test Method									NHK NRU Test Method								
		ECBC			FAL			IIVS			ECBC			FAL			IIVS		
		IC_{50}[1] µg/mL	SD[2]	N	IC_{50}[1] µg/mL	SD[2]	N	IC_{50}[1] µg/mL	SD[2]	N	IC_{50}[1] µg/mL	SD[2]	N	IC_{50}[1] µg/mL	SD[2]	N	IC_{50}[1] µg/mL	SD[2]	N
Sodium selenate	Sodium, selenium compound	12.7	1.62	3	54.2	10.4	3	36.5	5.23	3	7.47	0.861	3	16.1	9.55	3	10.0	1.33	3
Phase III																			
Acetaminophen	Amide	40.8	9.12	3	66.2	23.0	3	43.4	11.4	3	558	80.7	3	447	83.7	3	571	79.0	3
Acetonitrile	Nitrile	6433	129	3	9690	5634	3	9330	1217	3	10868	7824	4	10153	1960	4	9290	413	3
Acetylsalicylic acid	Carboxylic Acid	646	61.5	3	1234	298	3	401	62.0	3	631	19.9	3	694	98.3	3	514	79.1	3
5-Aminosalicylic acid	Carboxylic Acid	1467	203	3	2070	334	3	1557	179	3	29.9	6.52	3	78.2	42.3	3	48.8	7.90	3
Amitriptyline HCl	Polycyclic	6.03	1.38	3	7.86	2.20	3	7.81	1.38	3	10.8	3.34	3	7.57	5.43	3	10.9	1.04	3
Atropine sulfate	Heterocyclic	54.1	29.6	3	133	41.1	3	70.0	5.7	3	85.4	10.5	3	104	88.2	3	83.2	21.0	3
Boric acid	Boron compound	1497	484	3	3987	693	3	1202	581	3	440	138	3	517	378	3	464	11.0	3
Busulfan	Alcohol	40.4	19.3	3	321	180	3	43.7	1.77	3	253	68.2	3	268	193	3	313	37.2	3
Caffeine	Heterocyclic	133	13.3	3	157	81.7	3	191	14.4	3	817	256	3	591	186	3	574	7.81	3
Carbamazepine	Heterocyclic	83.0	12.0	3	152	56.9	3	91.8	11.0	3	66.1	8.40	3	253	325	3	63.9	5.27	3
Carbon tetrachloride	Halogenated hydrocarbon	NA	NA	-	NA	NA	-	NA	NA	-	NA	NA	-	NA	NA	-	NA	NA	-
Chloral hydrate	Alcohol	151	15.6	3	241	25.1	3	170	19.9	3	140	34.2	3	159	50.1	3	112	1.73	3
Citric acid	Carboxylic acid	473	138	3	1148	143	4	865	160	3	526	82.4	3	312	51.6	4	433	22.3	3

Table 5-3   3T3 and NHK NRU Test Method Summary IC$_{50}$ Data from the Laboratories

Substance	Chemical Class[4]	3T3 NRU Test Method									NHK NRU Test Method								
		ECBC			FAL			IIVS			ECBC			FAL			IIVS		
		IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N
Cupric sulfate pentahydrate	Sulfur compound	82.7	3.18	3	123	54.0	4	5.72	1.75	3	190	19.6	3	195	12.5	3	207	7.09	3
Cycloheximide	Heterocyclic	0.125	0.057	3	0.647	0.451	3	0.109	0.025	3	0.053	0.012	3	0.120	0.094	3	0.071	0.013	3
Dibutyl phthalate	Carboxylic acid	23.5	3.98	3	191	94.5	4	20.7	1.37	3	28.3	7.64	3	47.4	34.3	3	22.0	1.32	3
Dichlorvos	Organophos- phorous	9.83	3.42	3	32.8	2.07	3	18.3	2.09	3	8.56	2.28	3	12.4	3.74	3	12.2	0.416	3
Diethyl phthalate	Carboxylic acid	85.5	29.0	3	147	37.8	3	106	25.3	3	174	14.4	3	71.5	67.3	3	189	33.1	3
Digoxin	Polycyclic	351	137	3	892	319	3	317	67.9	2	0.0054	0.0007	3	0.0001	0.00002	3	0.0040	0.0003	3
Dimethyl- formamide	Amide	5343	515	3	5483	517	3	4900	183	3	9353	155	3	7817	100	3	6397	202	3
Diquat dibromide monohydrate	Heterocyclic	3.87	0.887	3	36.1	35.5	3	5.39	1.36	3	3.59	0.825	3	6.77	3.73	4	3.84	0.313	3
Disulfoton	Organophos- phorous compound	137	74.9	3	11200	NA	1	60.4	52.5	3	140	27.0	3	808	213	3	186	59.2	3
Endosulfan	Heterocyclic	5.27	3.01	3	15.2	11.9	4	3.61	1.53	3	3.44	0.573	3	1.42	0.701	4	2.19	0.437	3
Epinephrine bitartrate	Alcohol	51.5	6.16	3	63.4	6.63	3	63.4	1.91	3	115	10.8	3	81.7	28.4	3	75.0	12.2	3
Ethanol	Alcohol	5360	1754	3	8420	1205	3	6413	345	3	8290	390	3	12013	2286	3	10250	867	3
Fenpropathrin	Hydrocarbon	22.6	2.41	3	42.4	26.8	4	16.7	2.03	3	3.73	1.01	3	2.23	0.616	3	1.82	0.310	3
Gibberellic acid	Hydrocarbon	8027	908	3	NA	NA	-	7657	745	3	2850	402	3	2940	276	3	2807	121	3

Table 5-3   **3T3 and NHK NRU Test Method Summary IC$_{50}$ Data from the Laboratories**

Substance	Chemical Class[4]	3T3 NRU Test Method									NHK NRU Test Method								
		ECBC			FAL			IIVS			ECBC			FAL			IIVS		
		IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N
Glutethimide	Heterocyclic	167	7.00	3	284	20.7	3	125	9.25	4	187	64.3	3	170	24.1	3	176	27.5	3
Glycerol	Alcohol	20000	2987	3	38878	28238	4	27833	10882	3	34267	15399	3	18023	8334	3	29033	4596	3
Haloperidol	Ketone	5.32	0.649	3	7.99	0.655	3	5.47	0.654	3	3.69	1.01	3	3.72	1.81	3	3.29	1.15	3
Hexachlorophene	Cyclic hydrocarbon	5.02	2.41	3	5.35	1.75	3	3.06	0.289	3	0.027	0.004	3	0.046	0.020	3	0.021	0.002	3
Lactic acid	Carboxylic acid	2943	315	3	3487	561	3	2790	259	3	1290	52.9	3	1320	60.8	3	1313	138	3
Lindane	Halogenated hydrocarbon	125	119	3	266	94.8	4	90.4	111	5	19.1	3.14	3	23.2	7.09	3	15.6	2.40	3
Meprobamate	Carboxylic acid	353	49.7	3	877	128	4	386	9.02	3	761	116	3	163	189	3	624	84.2	3
Mercury II chloride	Mercury compound	3.45	0.177	3	5.99	1.87	3	3.51	0.120	3	6.87	1.04	3	5.40	1.02	3	5.35	0.090	3
Methanol	Alcohol	NA	NA	-	NA	NA	-	NA	NA	-	NA	NA	-	1133	213	3	2100	226	3
Nicotine	Heterocyclic	272	65.3	3	412	136	3	450	54.7	3	94.3	24.7	3	134	78.4	3	112	27.7	3
Paraquat	Heterocyclic	21.3	7.29	3	24.9	16.5	3	23.7	15.2	3	48.3	6.03	3	96.6	37.2	3	53.4	5.52	3
Parathion	Organophos-phorous compound	22.7	12.1	3	141	98.7	4	22.0	4.94	3	34.0	10.0	3	31.2	11.9	3	29.0	8.34	3
Phenobarbital	Heterocyclic	634	134	3	726	255	3	476	111	4	693	180	3	360	95.5	3	381	69.9	3
Phenol	Phenol	50.2	10.9	3	104	24.8	3	58.1	6.78	3	59.1	21.4	3	93.2	5.97	3	80.8	5.12	3

Table 5-3 **3T3 and NHK NRU Test Method Summary IC$_{50}$ Data from the Laboratories**

| Substance | Chemical Class[4] | 3T3 NRU Test Method | | | | | | | | | NHK NRU Test Method | | | | | | | | | |
| | | ECBC | | | FAL | | | IIVS | | | ECBC | | | FAL | | | IIVS | | |
		IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N
Phenylthiourea	Sulfur compound	30.1	19.8	3	239	65.8	3	89.0	21.9	3	363	58.0	3	401	83.6	3	272	71.7	3
Physostigmine	Carboxylic acid	28.2	14.9	3	37.8	1.93	3	20.4	6.71	4	164	5.51	3	212	238	3	139	8.74	3
Potassium cyanide	Potassium, nitrogen compound	15.3	3.76	3	159	81.9	3	18.9	0.950	3	29.3	6.90	3	89.0	100	3	16.9	2.21	3
Procainamide HCl	Amide	400	15.3	3	431	4.73	3	497	39.3	3	1480	200	3	1787	221	3	2027	229	3
Propylparaben	Carboxylic acid	20.9	3.33	3	51.8	14.8	3	17.1	2.10	3	18.1	2.42	3	18.6	2.84	3	13.8	1.21	3
Sodium arsenite	Arsenical	0.496	0.028	3	1.44	0.819	3	0.683	0.117	3	0.790	0.248	3	0.336	0.187	3	0.470	0.066	3
Sodium chloride	Sodium, chlorine compound	4790	233	3	4625	611	4	4877	457	3	3583	263	3	1118	1388	3	3470	300	3
Sodium dichromate dihydrate	Sodium, chromium compound	0.603	0.087	3	0.657	0.244	3	0.547	0.092	3	0.784	0.113	3	0.851	0.302	4	0.576	0.100	3
Sodium hypochlorite	Sodium, oxygen, chlorine compound	823	108	3	805	367	3	2005	872	4	1863	581	3	1243	576	3	1633	180	3
Sodium oxalate	Carboxylic acid	42.0	17.3	3	31.0	8.66	3	49.5	26.3	4	355	54.9	3	350	147	4	360	94.6	3
Strychnine	Heterocyclic	389	80.9	3	124	20.3	3	83.5	5.35	3	100	76.6	4	52.5	28.0	3	55.1	3.43	3
Thallium I sulfate	Metal	2.81	0.671	3	13.4	10.4	4	6.27	1.75	3	0.198	0.100	3	0.153	0.031	3	0.127	0.020	3

Table 5-3   **3T3 and NHK NRU Test Method Summary IC$_{50}$ Data from the Laboratories**

Substance	Chemical Class[4]	3T3 NRU Test Method									NHK NRU Test Method								
		ECBC			FAL			IIVS			ECBC			FAL			IIVS		
		IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N	IC$_{50}$[1] µg/mL	SD[2]	N
Trichloroacetic acid	Carboxylic acid	762	99.1	3	1220	72.1	3	801	114	3	348	63.5	3	541	150	3	394	50.8	3
1,1,1-Trichloro-ethane	Halogenated hydrocarbon	41100	NA	1	21250	2357	3	9827	180	3	8137	591	3	NA	NA	-	NA	NA	-
Triethylene-melamine	Triazine	0.086	0.009	3	1.45	0.265	3	0.169	0.049	3	1.69	0.950	3	2.03	0.471	3	2.13	0.480	3
Triphenyltin hydroxide	Organo-metallic compound	0.026	0.004	3	0.026	0.021	3	0.015	0.008	3	0.021	0.007	3	0.007	0.007	3	0.011	0.003	3
Valproic acid	Carboxylic acid	547	67.1	3	1807	175	3	574	NA	1	468	116	3	702	160	3	430	71.5	3
Verapamil HCl	Amine	32.2	5.82	3	34.6	1.72	3	38.9	4.20	3	60.5	13.6	3	79.4	33.9	3	66.2	5.57	3
Xylene	Cyclic hydrocarbon	NA	NA	-	NA	NA	-	724	87.1	3	NA	NA	-	NA	NA	-	486	185	3

[1]Arithmetic mean

[2]Standard deviation

[3]Data are slightly different from that summarized in **Table 5-2** for Phase Ia. These data represent the acceptable tests after implementation of the R^2 acceptance criterion, while the data in **Table 5-2** represent acceptable tests prior to the implementation of the criterion.

[4]Chemical class assigned is based on the classification of the National Library of Medicine's Medical Subject Heading (MeSH), http://www.nlm.nih.gov/mesh/meshhome.html

NA = not available; IC$_{50}$ values could not be generated (see footnotes in **Appendix J**)

546   **Figure 5-1   3T3 NRU IC$_{50}$ Values by Reference Substance and Laboratory**

547   **(Substances are grouped from lowest mean IC$_{50}$ value (aminopterin) to highest mean IC$_{50}$ value (ethylene**

548   **glycol).**

549   a

3T3: IC$_{50}$ Values by Chemical and Laboratory

550
551   b

3T3: IC$_{50}$ Values by Chemical and Laboratory

552

553
554   c

3T3: IC$_{50}$ Values by Chemical and Laboratory

555
556   d

3T3: IC$_{50}$ Values by Chemical and Laboratory

557
558
559
560

561
562   e

3T3: IC$_{50}$ Values by Chemical and Laboratory

563
564   f

3T3: IC$_{50}$ Values by Chemical and Laboratory
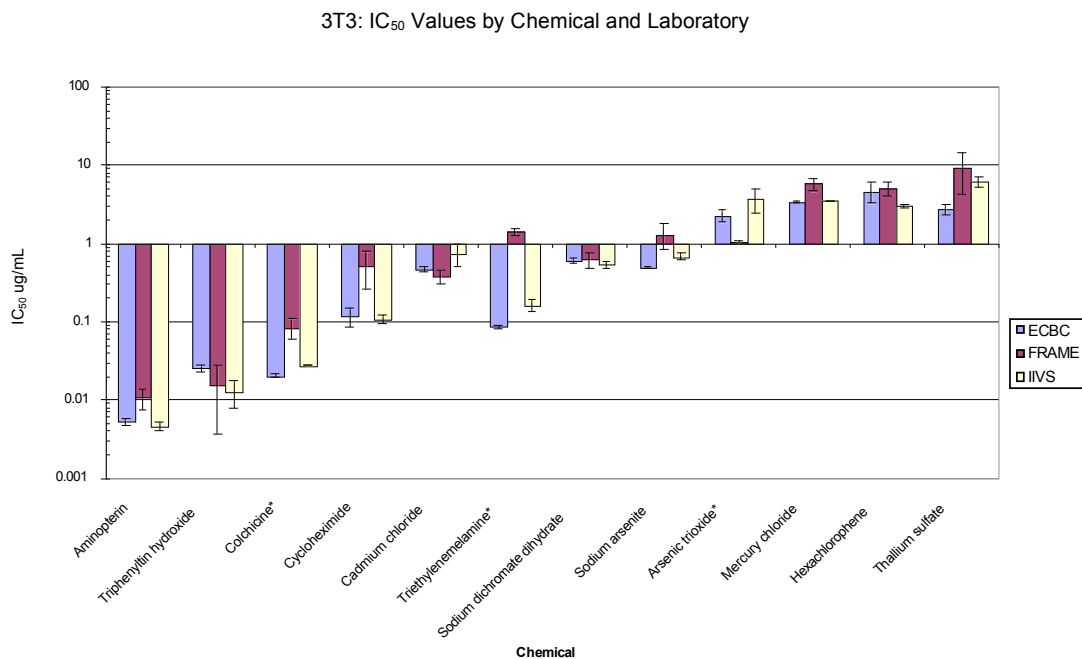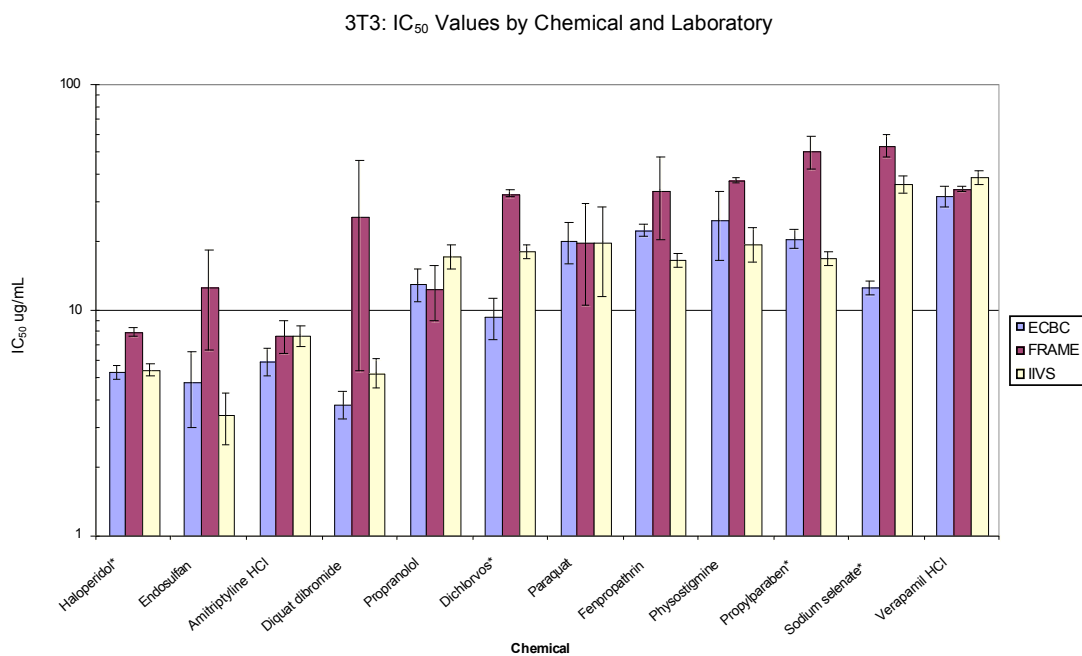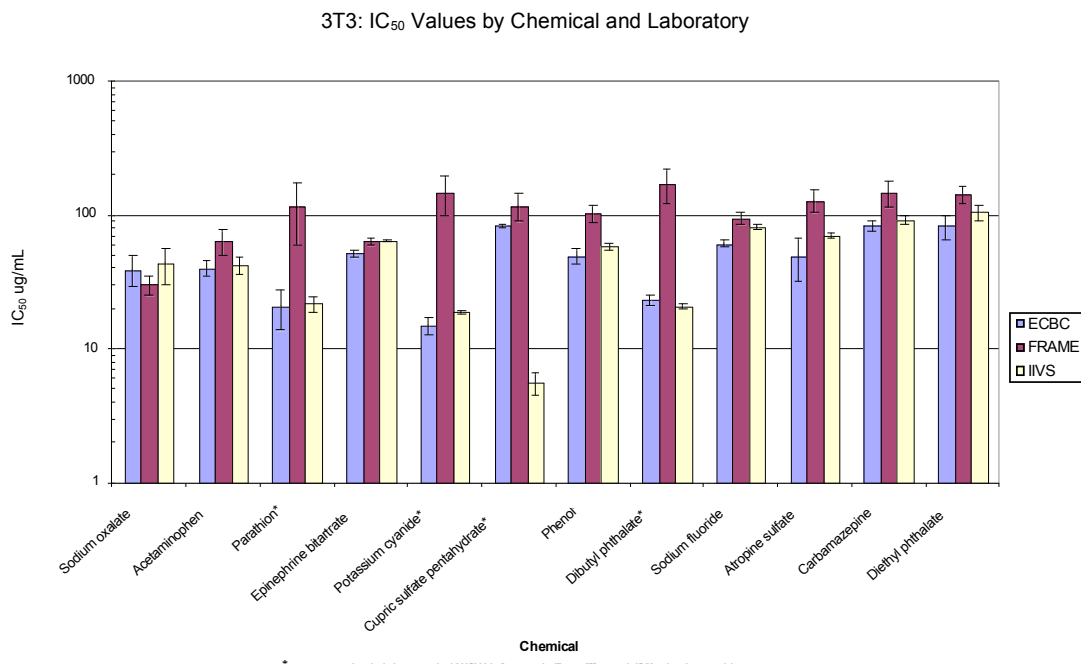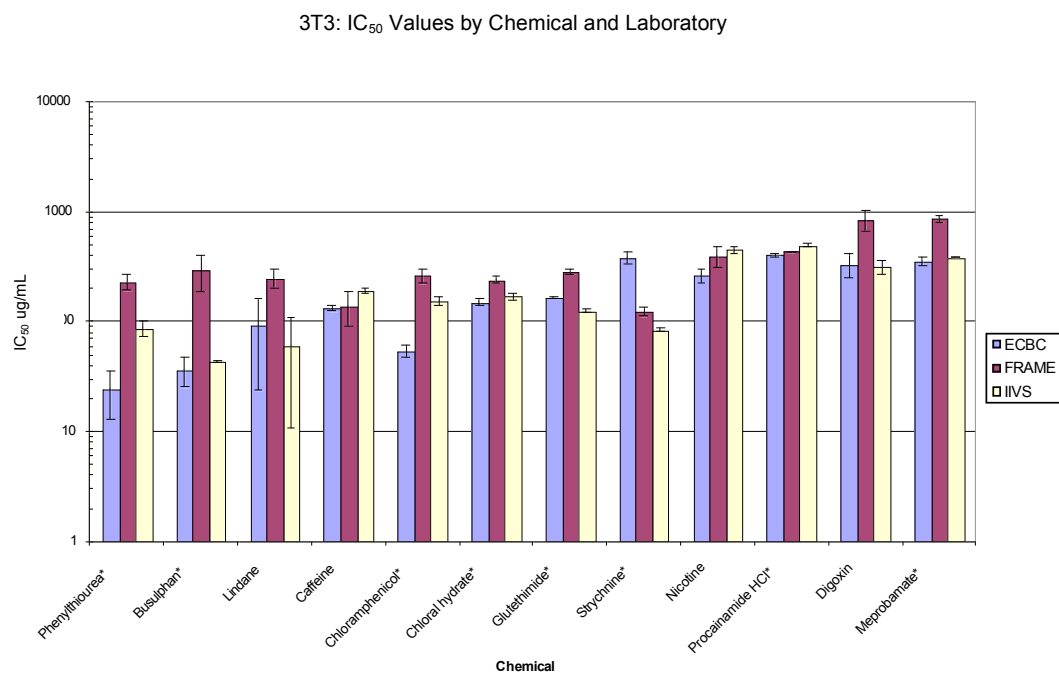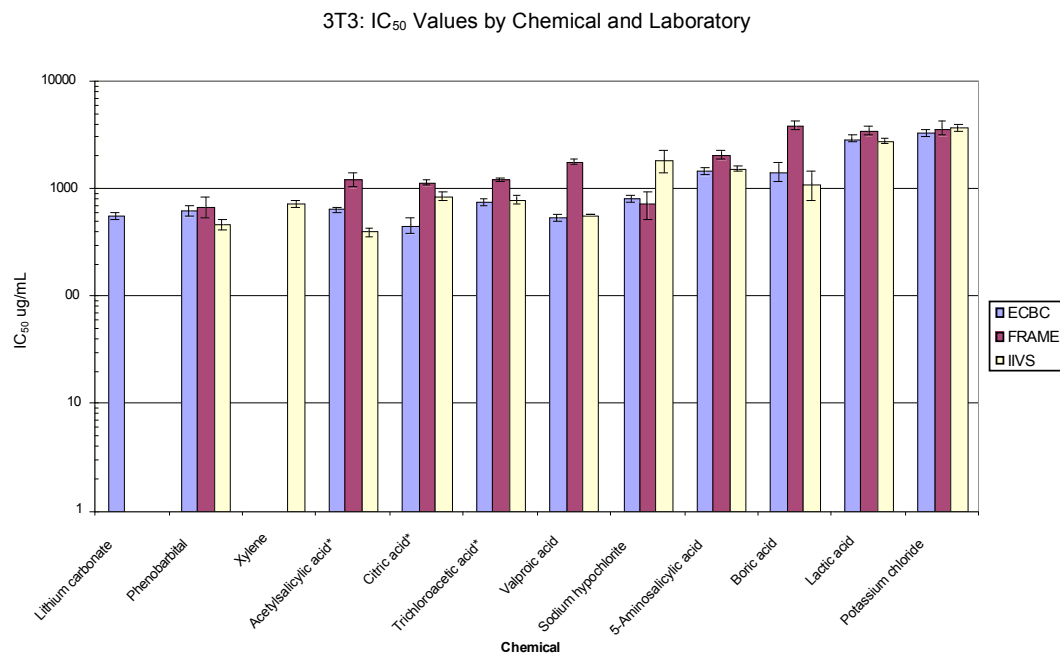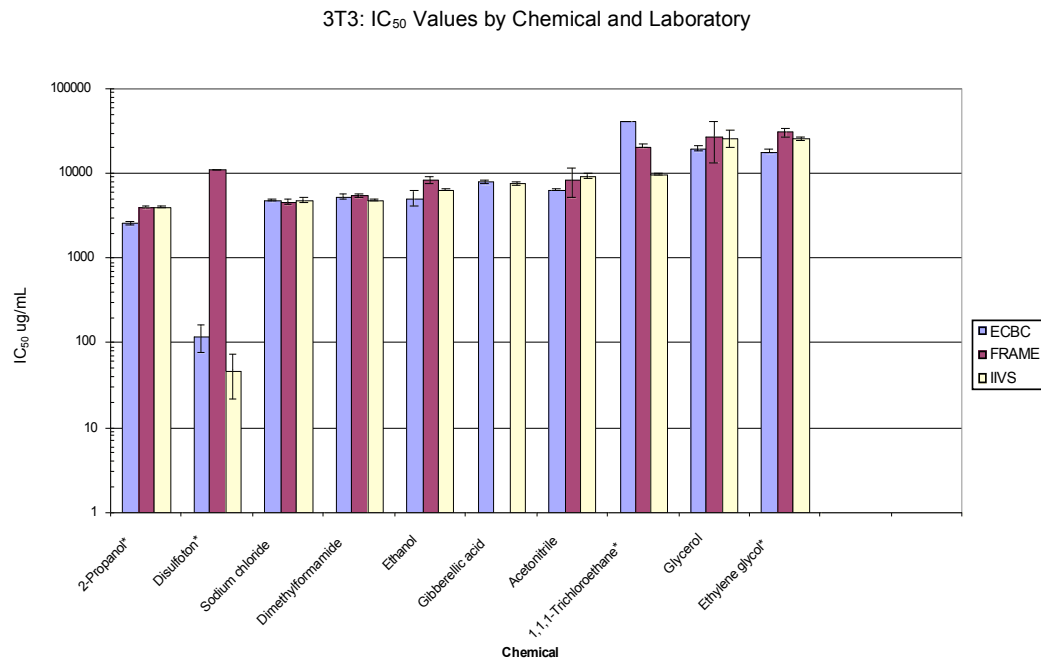
565
566   *Represents a chemical where the standard ANOVA indicates a significant difference in IC$_{50}$ values
567   between laboratories. Bars represent mean IC$_{50}$ from each laboratory in μg/mL. Log IC$_{50}$ values used
568   to allow multiple data sets on each graph. Error bars show the standard deviation.
569

569 **Figure 5-2   NHK NRU IC$_{50}$ Values by Reference Substance and Laboratory**
570 **(Substances are grouped from lowest mean IC$_{50}$ value (digoxin) to the highest mean IC$_{50}$ value (ethylene**

571 **glycol).**

572 a

NHK: IC$_{50}$ Values by Chemical and Laboratory

573
574 b

NHK: IC$_{50}$ Values by Chemical and Laboratory

575

576 c

NHK: IC$_{50}$ Values by Chemical and Laboratory

577
578 d

NHK: IC$_{50}$ Values by Chemical and Laboratory
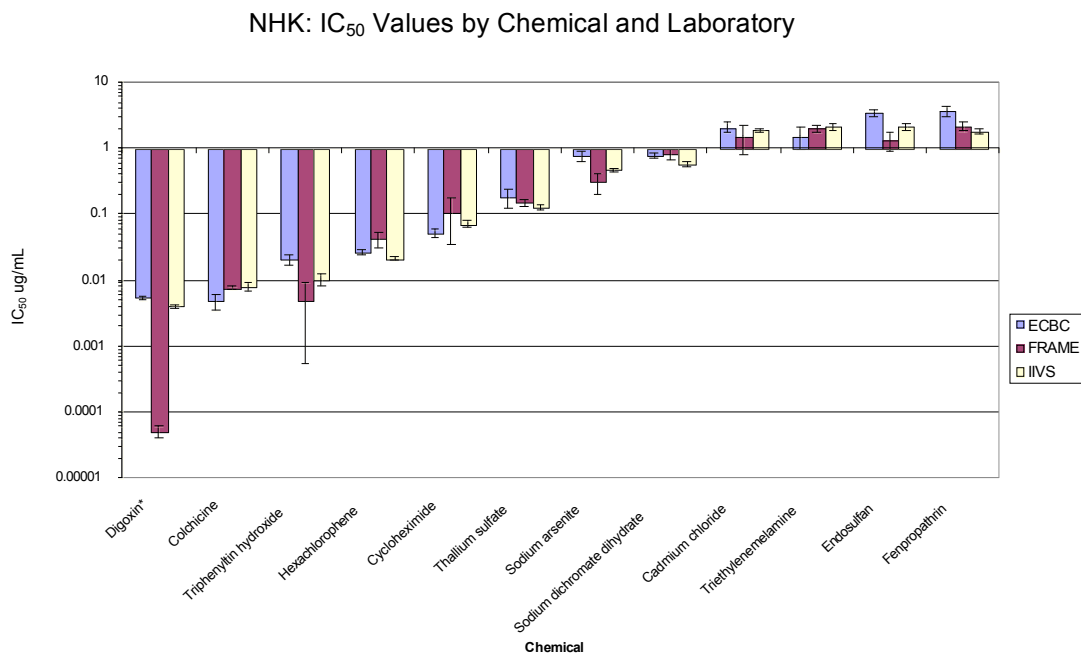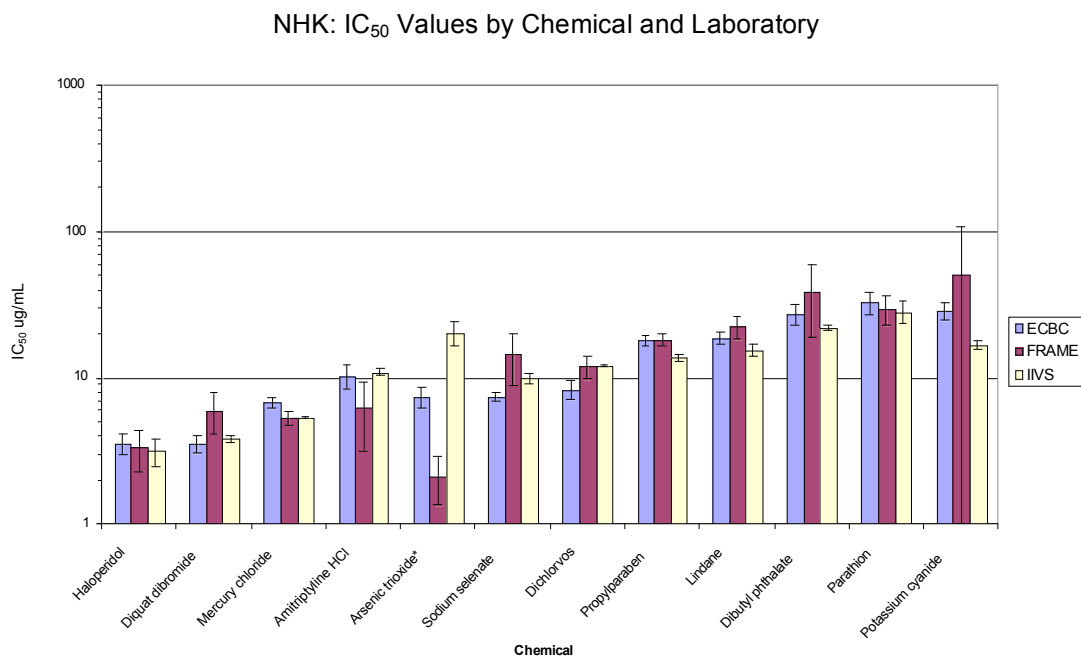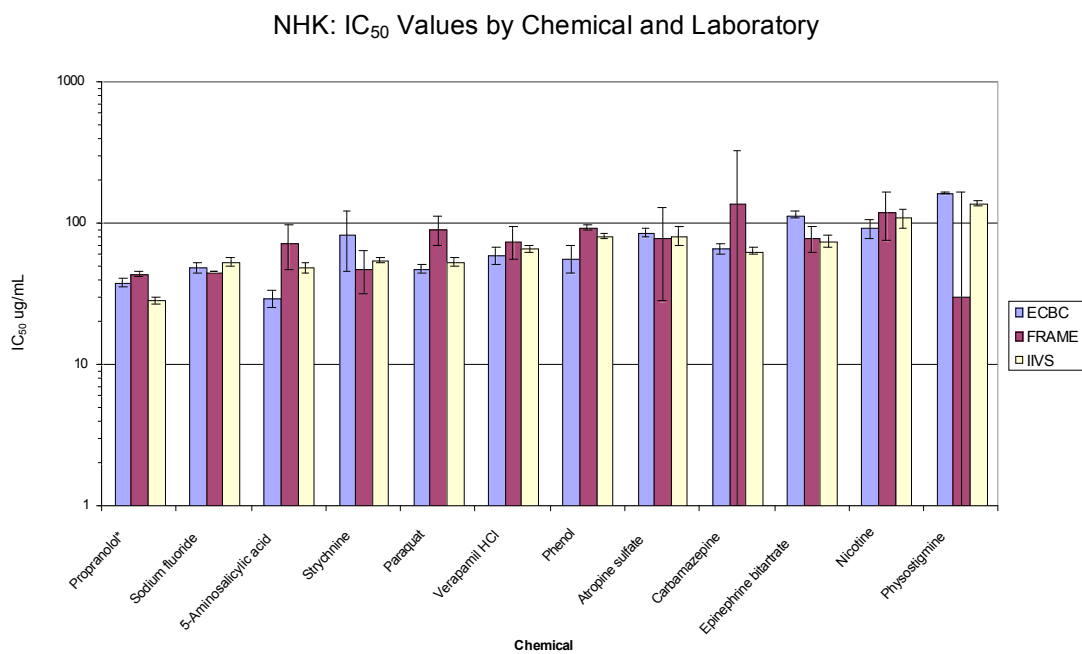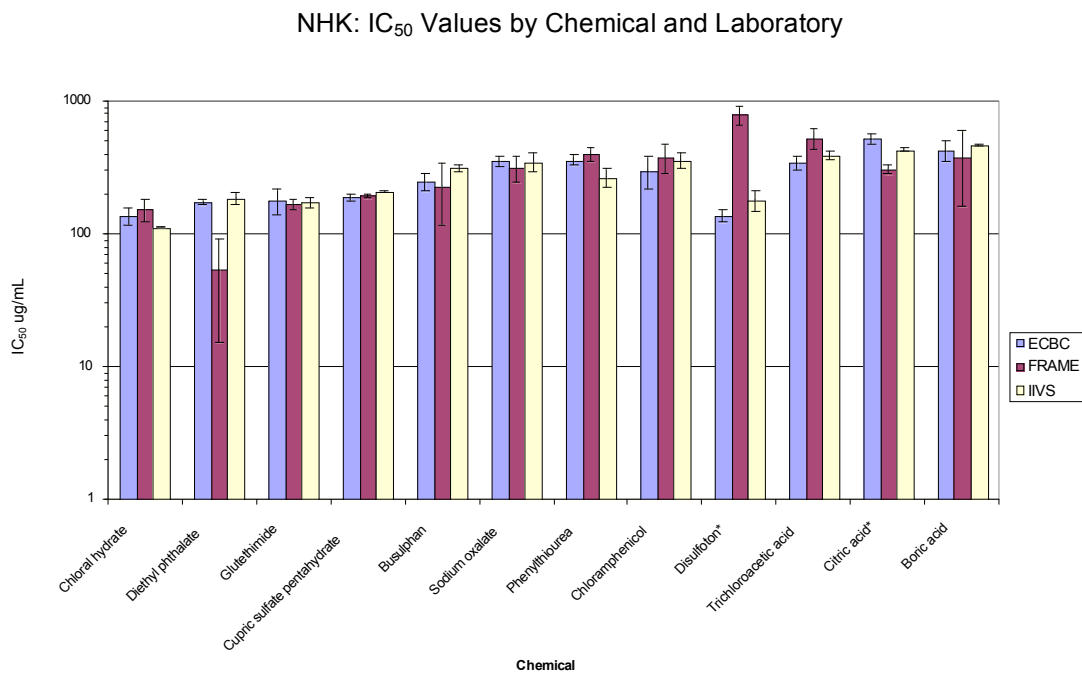
579
580
581
582
583

584   e

NHK: IC$_{50}$ Values by Chemical and Laboratory

585
586   f
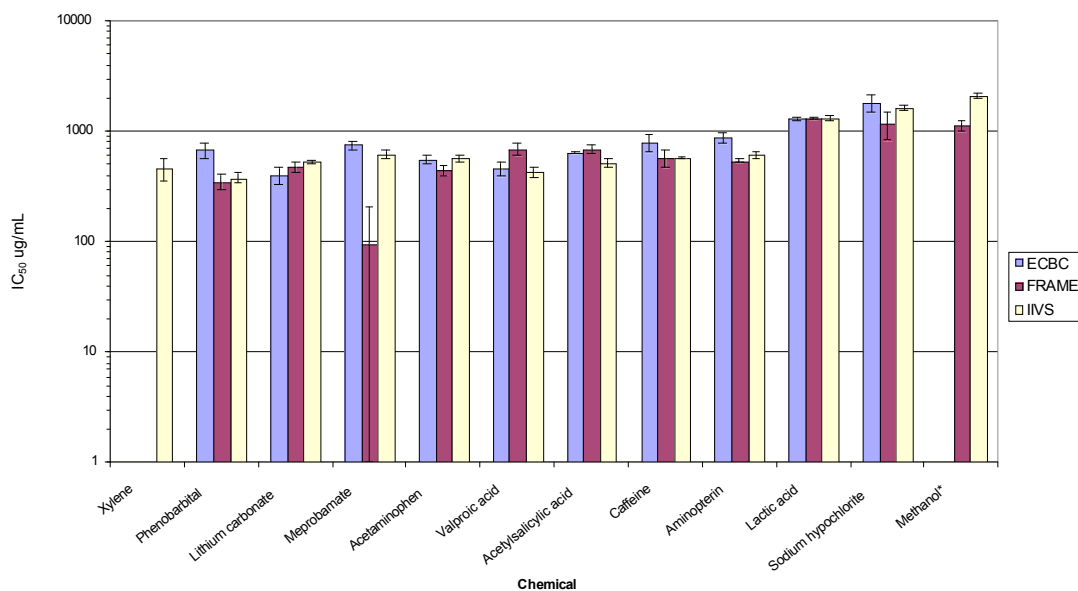
NHK: IC$_{50}$ Values by Chemical and Laboratory
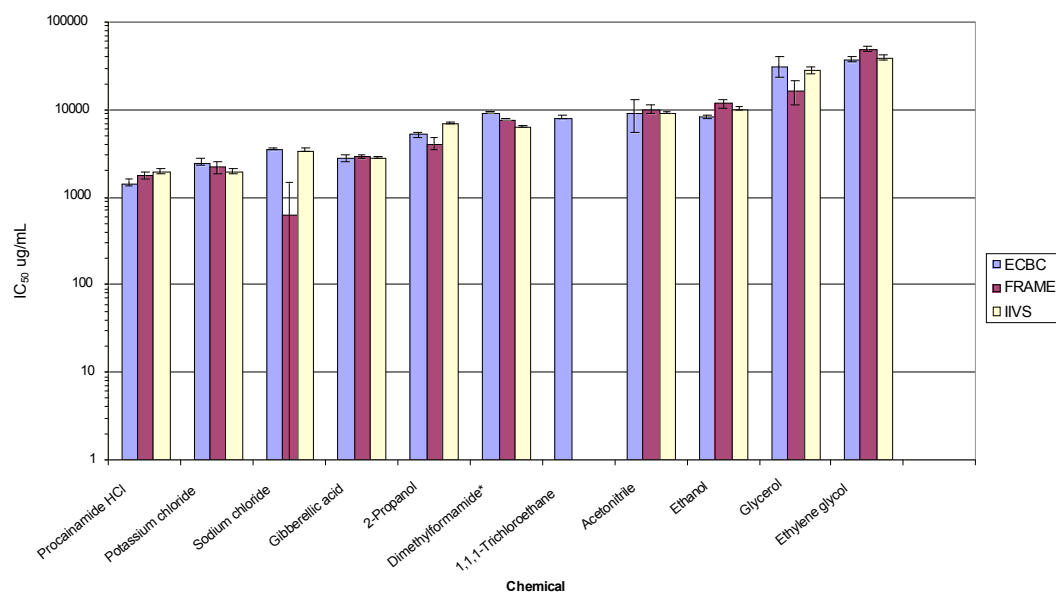
587
588   *Represents a chemical where the standard ANOVA indicates a significant difference in IC$_{50}$ values
589   between laboratories. Bars represent mean IC$_{50}$ from each laboratory in μg/mL. Log IC$_{50}$ values used
590   to allow multiple data sets on each graph. Error bars show the standard deviation.

591

Table 5-4   Comparison of 3T3 and NHK IC$_{50}$ Geometric Means

Reference Substance	3T3 NRU Test Method Geometric Mean[1] IC$_{50}$ (μg/mL)	NHK NRU Test Method Geometric Mean[1] IC$_{50}$ (μg/mL)	Difference (Orders of Magnitude)
Carbon tetrachloride	NA	NA	NA
Methanol	NA	1529[b]	NA
Aminopterin	0.006	669	5
Triphenyltin hydroxide	0.017	0.010	0
Colchicine	0.034	0.007	1
Cycloheximide	0.187	0.073	1
Triethylenemelamine	0.272	1.85	1
Cadmium II chloride	0.518	1.84	1
Sodium dichromate dihydrate	0.587	0.721	0
Sodium arsenite	0.759	0.477	0
Arsenic trioxide	1.96	5.26	0
Mercury II chloride	4.12	5.80	0
Hexachlorophene	4.19	0.029	2
Thallium I sulfate	5.74	0.152	1
Haloperidol	6.13	3.36	0
Endosulfan	6.35	2.13	0
Amitriptyline HCl	7.05	8.96	0
Diquat dibromide monohydrate	8.04	4.48	0
Propranolol	13.9	35.3	0
Dichlorvos	17.7	10.7	0
Paraquat	20.1	61.6	0
Fenpropathrin	24.2	2.43	1
Physostigmine	25.8	88.5	0
Propylparaben	26.1	16.6	0
Sodium selenate	29.0	10.2	0
Potassium cyanide	34.6	29.0	1
Verapamil HCl	34.9	66.5	0
Parathion	37.4	30.3	0
Sodium oxalate	37.7	337	1
Sodium lauryl sulfate (SLS)*	*41.7*	*3.99*	*1*
Cupric sulfate pentahydrate	42.1	197	1
Acetaminophen	47.7	518	1
Dibutyl phthalate	49.7	28.7	0
Epinephrine bitartrate	59.0	87.4	0
Phenol	66.3	75.0	1
Atropine sulfate	76.0	81.8	0
Busulfan	77.7	260	1
Sodium I fluoride	78	49.8	0
Phenylthiourea	79.0	336	1
Carbamazepine	103	83.2	1
Diethyl phthalate	107	120	0
Lindane	108	18.7	1
Chloramphenicol	128	348	0
Disulfoton	133	270	0
Caffeine	153	638	0
Strychnine	158	62.5	1
Glutethimide	174	174	0

Table 5-4   Comparison of 3T3 and NHK IC$_{50}$ Geometric Means

Reference Substance	3T3 NRU Test Method Geometric Mean[1] IC$_{50}$ (μg/mL)	NHK NRU Test Method Geometric Mean[1] IC$_{50}$ (μg/mL)	Difference (Orders of Magnitude)
Chloral hydrate	183	133	0
Nicotine	361	107	0
Procainamide HCl	441	1741	1
Digoxin	466	0.001	5
Meprobamate	519	357	0
Lithium I carbonate	562[a]	468	0
Phenobarbital	573	448	0
Acetylsalicylic acid	676	605	0
Xylene	721[a]	466[a]	0
Citric acid	796	400	0
Trichloroacetic acid	902	413	0
Valproic acid	916	512	0
Sodium hypochlorite	1103	1502	0
5-Aminosalicylic acid	1667	46.7	2
Boric acid	1850	421	1
Lactic acid	3044	1304	0
Potassium I chloride	3551	2237	0
2-Propanol	3618	5364	0
Sodium chloride	4730	1997	0
Dimethylformamide	5224	7760	0
Ethanol	6523	10018	1
Gibberellic acid	7810[b]	2856	0
Acetonitrile	7951	9528	0
1,1,1-Trichloroethane	17248	8122[a]	1
Ethylene glycol	24317	41852	0
Glycerol	24655	24730	0

Table sorted by 3T3 IC$_{50}$ values
[1]Laboratories combined; use of a geometric mean for the IC$_{50}$ values in **Table 5-4** is consistent with the approach used for the RC millimole regression to obtain a single IC$_{50}$ from multiple IC$_{50}$ values (Halle 1998).
[a]Data available from only one laboratory
[b]Data available from only two laboratories
*Positive control (SLS) values (met acceptance criteria) from all test phases: N = 293 (3T3); N = 281 (NHK)
NA = not available

602   Two chemicals, digoxin and aminopterin, have IC_{50} values that differ by five orders of

603   magnitude between the two cell types. Digoxin was much more toxic to the NHK cells and

604   aminopterin was more toxic to the 3T3 cells. Hexachlorophene and 5-aminosalicylic acid

605   IC_{50} values were different by two orders of magnitude and both were more toxic to the NHK

606   cells than the 3T3 cells. The positive control (SLS) values for the two cell types differed by

607   an order of magnitude (41.7 μg/mL for 3T3; 3.99 μg/mL for NHK). Of the IC_{50} reference

608   substance values. 94.5% for both cell types were within at least 2 orders of magnitude of

609   each other. **Table 5-5** illustrates the comparisons of the IC_{50} values.

610

611   **Table 5-5**   **Difference in 3T3 and NHK NRU IC_{50} Values as Orders of Magnitude**

Difference (Orders of Magnitude)	Percentage of Reference Substances
0	63.9% (46/72)
1	27.8% (20/72)
2	2.8% (2/72)
3	0
4	0
5	2.8% (2/72)
NA	2.8% (2/72)

612

613

614

615

615   **Figure 5-3**   **RC IC_{50} Values vs 3T3 NRU IC_{50} Values for the 58 Common Chemicals**

616–628

$r_s = 0.93$
$n = 56$

3T3 NRU IC_{50} (mM)

Registry of Cytotoxicity IC_{50} (mM)
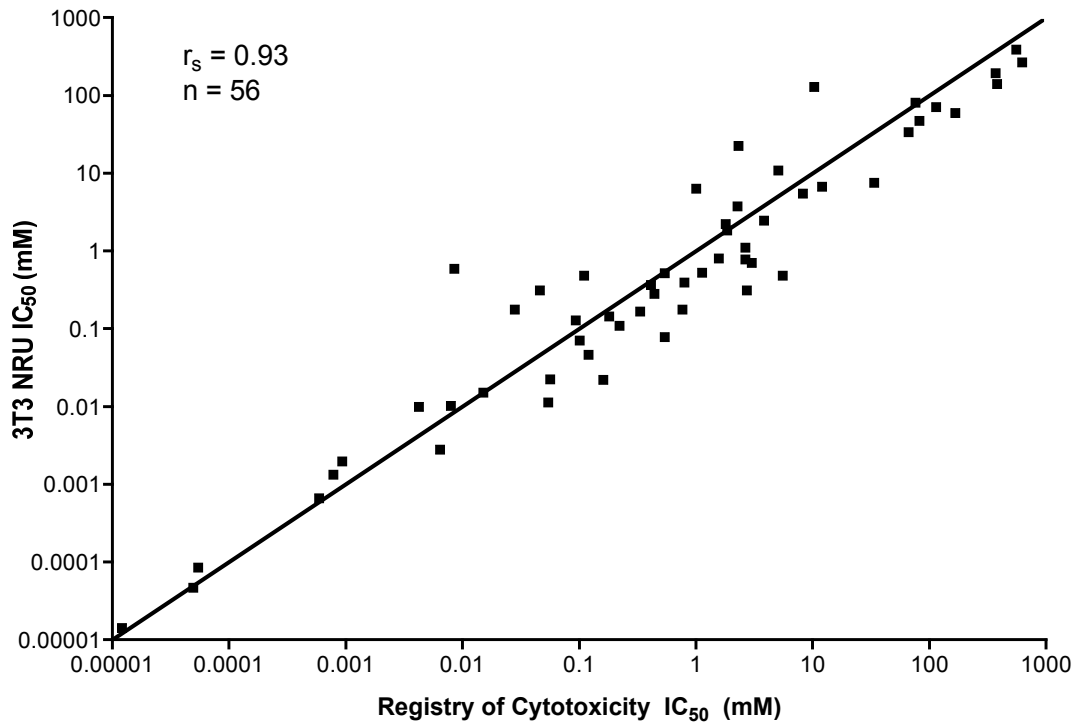
631   **Figure 5-4**   **RC IC_{50} Values vs NHK NRU IC_{50} Values for the 58 Common Chemicals**

$r_s = 0.86$

NHK NRU IC_{50} (mM)

Registry of Cytotoxicity IC_{50} (mM)
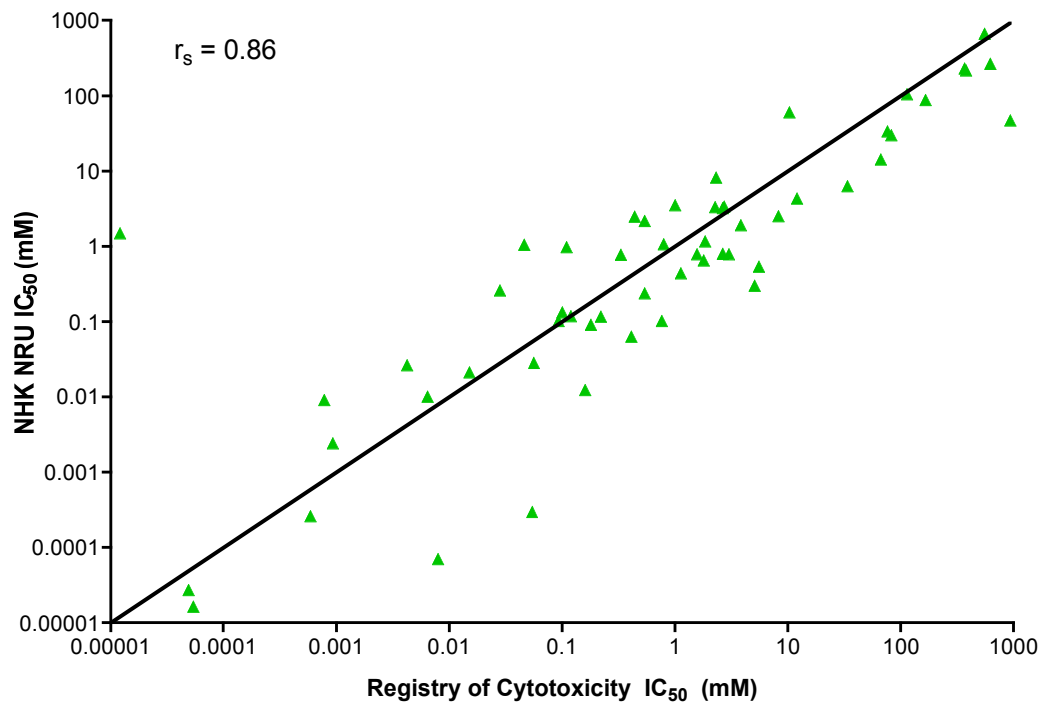
633 **5.5   Coded Reference Substances and GLP Guidelines**

634

635 5.5.1   Coded Reference Substances

636 BioReliance acquired 73 high purity chemicals (72 reference substances and one positive

637 control chemical, at 99% or greater purity when economically feasible) from reputable

638 commercial sources (see **Appendix F**). BioReliance randomly coded each reference

639 substance with a unique identification number when repackaging into multiple smaller units.

640 These units were given an additional code unique for the respective cytotoxicity laboratories

641 so that substances could be provided in a blinded fashion (see **Section 3.6** for distribution

642 procedures). The reference substances were packaged and shipped such that their identities

643 were concealed; however, all laboratories knew the identity of the positive control. The SMT

644 revealed the reference substance codes for each phase after all laboratories had submitted

645 their data and reports. Periodically, laboratories required additional aliquots of reference

646 substance and BioReliance provided these aliquots from the original stock of reference

647 substance in the same manner that the original aliquots were provided.

648

649 5.5.2   Lot-to-Lot Consistency of Reference Substances

650 One lot of each substance was purchased and each laboratory received aliquots from this

651 same lot throughout the validation study. The substance suppliers provided certificates of

652 analysis for each lot along with other chemical, physical, and safety information concerning

653 the substance (e.g., MSDS documents).

654

655 5.5.3   Adherence to GLP Guidelines

656 BioReliance, ECBC, and IIVS, followed GLP procedures for all testing with the exception of

657 tests designed to resolve technical challenges (e.g., formation of NR crystals, use of film

658 plate sealers for volatile substances, slow growth of cells, etc.). These laboratories submitted

659 data to their respective quality assurance unit (as per GLP requirements) and copies of the

660 data were submitted to NICEATM. FAL followed most GLP guidelines, but their activities

661 did not include independent quality assurance reviews of laboratory procedures or

662 documentation. The Study Director for the FAL performed all data reviews and provided

663   copies to NICEATM. Hard copy printouts of all data as well as electronic versions are

664   available at NICEATM.

665

666   **5.6   Study Timeline and NICEATM/ECVAM Study Participatory Laboratories**

667

668   5.6.1   Statement of Work (SOW) and Protocols

669   The SMT provided the laboratories with an SOW prior to initiation of testing (see **Appendix**

670   **G**) and proposed dates for completion of various aspects of the study (e.g., transfer of data,

671   provision of reports, etc.). The SOW for the cytotoxicity laboratories defined the following:

672   •   project objectives

673   •   management and key personnel

674   •   required facilities, equipment, and supplies

675   •   quality assurance requirements

676   •   test phases and schedules

677   •   products (e.g., reports) required

678   •   report preparation

679

680   The SOW for BioReliance contained all of the above and included requirements for:

681   •   reference substance acquisition, preparation, and distribution

682   •   solubility testing

683

684   The SMT, in consultation with the laboratories, prepared Test Method Protocols for each

685   phase of the study. Cytotoxicity testing for each phase (in each laboratory) was initiated

686   when the SMT received a signed protocol specific for that phase from the Study Director.

687   Solubility testing for Phases I and II was performed prior to cytotoxicity testing for those

688   phases while solubility testing for the Phase III substances was performed throughout Phases

689   II and III.

690

691

692

693

694   5.6.2   Study Timeline

695   The actual timeline achieved in the study is shown in **Table 5-6**. The SMT eased the original

696   timeline presented in the SOWs due to various factors (e.g., protocol revisions, side studies,

697   acquisition of medium, etc.).

698

699   **Table 5-6   Validation Study Timetable**

700

	BioReliance	ECBC	FAL	IIVS
Receipt of SOW	Jun 2002	Jun 2002	Jun 2002	Jun 2002
Procurement of Chemicals	Jul 2002 - Jan 2003	NA	NA	NA
Solubility Testing	Jul 2002 - Jan 2003	Sep 2004	Dec 2003	Jan 2004
Distribution of Reference Substances Phase Ia Phase Ib Phase II Phase III	Jul 2002 Sep 2002 Nov 2002 Feb - Mar 2003	NA	NA	NA
Initiation of Phase Ia	NA	Aug 2002	Aug 2002	Aug 2002
Completion of Phase Ia	NA	Nov 2002	Nov 2002	Oct 2002
Initiation of Phase Ib	NA	Dec 2002	Dec 2002	Dec 2002
Completion of Phase Ib	NA	May 2003	May 2003	May 2003
Initiation of Phase II	NA	Jun 2003	Jun 2003	Jun 2003
Completion of Phase II	NA	Nov 2003	Nov 2003	Nov 2003
Initiation of Phase III	NA	Dec 2003	Dec 2003	Dec 2003
Completion of Phase III	NA	Dec 2004	Dec 2004	Jan 2005

701   NA- not applicable; SOW = BioReliance distributed reference substances; ECBC, FAL, AND IIVS tested the
702   reference substances

703

704   5.6.3   Participatory Laboratories

705

706   BioReliance Corporation

707   14920 Broschart Road

708   Rockville, Maryland 20850-3349

709   Study Director: Dr. Martin Wenk

710

711

712

713

714   U.S. Army Edgewood Chemical & Biological Center (ECBC)

715   Molecular Engineering Team

716   Aberdeen Proving Ground, MD 21010

717   Study Director: Dr. Cheng Cao

718

719   Institute for In Vitro Sciences (IIVS)

720   21 Firstfield Road Suite 220

721   Gaithersburg, MD 20878

722   Study Director: Mr. Hans Raabe

723

724   FRAME (Fund for the Replacement of Animals in Medical Experiments)

725   Alternatives Laboratory (FAL)

726   Queens Medical Centre

727   University of Nottingham

728   Nottingham NG7 2UH

729   United Kingdom

730   Study Director: Dr. Richard Clothier

731

732   **5.7   Availability of Data**

733

734   All data were submitted and provided to the SMT via electronic files and paper copies. The

735   laboratories also maintained copies of all raw data and the electronic files.

736

737   **5.8   Solubility Test Results**

738

739   This study evaluated a solubility protocol (see **Section 2-7** and **Appendix B-3**) designed to

740   identify the solvent that would provide the highest soluble concentration of a reference

741   substance for *in vitro* testing. Each laboratory performed a solubility test on all reference

742   substances. To avoid the use of different solvents by the laboratories when testing the same

743   substance, the SMT assigned the solvents used for *in vitro* testing (see **Table 6-9**). The

744   objectives of the solubility testing were to evaluate the utility and appropriateness of the

745　solubility protocol and to evaluate the concordance among laboratories in the solvent selected

746　for each of the 72 reference substances.

747

748　5.8.1　Solubility Data

749　BioReliance was the first laboratory to evaluate the solubility of the reference substances,

750　first in media, then in DMSO, and then in ETOH at 400 and 200 mg/mL. Based on this

751　experience, a solubility protocol for the *in vitro* laboratories was developed to test at lower

752　test article concentrations and to test with the various solvents at concentrations that would

753　be equivalent when applied to the cultures (see **Table 2-5**). The solubility flow chart (**Figure**

754　**2-7**) illustrates the tests for chemical solubility in medium, DMSO, and ETOH. **Table 5-7**

755　provides the solubility results in mg/mL.

756

Table 5-7   Solubility Results (data presented in mg/mL)

Reference Substance	BioReliance[1]				SMT[2] Selection	ECBC[3]				FAL[3]				IIVS[3]			
	3T3[4] Medium	NHK[5] Medium	DMSO	ETOH		3T3[4] Medium	NHK[5] Medium	DMSO	ETOH	3T3[4] Medium	NHK[5] Medium	DMSO	ETOH	3T3[4] Medium	NHK[5] Medium	DMSO	ETOH
Phase I																	
Arsenic III trioxide	0.25	0.05	< 2	< 2	Medium	*0.025[6]*	*0.025[6]*	*< 0.2*	*< 0.2*	*0.135[6]*	*0.135[6]*	*< 0.2*	*< 0.2*	*< 0.02[6]*	*< 0.02[6]*	*< 0.2*	*< 0.2*
Ethylene glycol	400	400	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Propranolol HCl	< 2	10	200	20	DMSO	0.2	2	200	NT	20	20	200	NT	20	2	NT	NT
Phase II																	
Aminopterin	2	2	NT	NT	DMSO	2.0	< 2	200	NT	< 2	2	200	NT	0.2	0.2	200	NT
Cadmium II chloride	< 2	< 2	200	< 200	DMSO	< 2	< 2	200	NT	< 2	< 2	200	NT	< 0.2	< 0.2	20	< 20
Chloramphenicol	2	2	400	< 200	DMSO	2.0	< 2	200	NT	< 2	< 2	200	NT	0.2	0.2	20	20
Colchicine	400	400	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Lithium I carbonate	0.25	10	< 2	NT	Medium	0.2	2.0	< 20	< 20	0.2	2	< 200	< 200	0.2	2	< 2	< 2
Potassium I chloride	200	200	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
2-Propanol	400	400	400	400	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Sodium I fluoride	20	20	< 200	< 200	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Sodium selenate	200	200	< 200	< 200	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Phase III																	
Acetaminophen	10	10	400	< 200	DMSO	2	2	NT	NT	2	2	NT	NT	< 2	< 2	200	NT
Acetonitrile	400	400	400	400	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Acetylsalicylic acid	10	10	400	200	DMSO	2	2	NT	NT	< 2	< 2	200	NT	2	2	NT	NT
5-Aminosalicylic acid	2	2	< 200	< 200	Medium	2	2	NT	NT	2	2	NT	NT	2	2	NT	NT
Amitriptyline HCl	200	200	NT	NT	DMSO	< 2	< 2	200	NT	< 2	< 2	200	NT	0.2	0.2	200	NT

Table 5-7   Solubility Results (data presented in mg/mL)

Reference Substance	BioReliance[1]				SMT[2] Selection	ECBC[3]				FAL[3]				IIVS[3]			
	3T3[4] Medium	NHK[5] Medium	DMSO	ETOH		3T3[4] Medium	NHK[5] Medium	DMSO	ETOH	3T3[4] Medium	NHK[5] Medium	DMSO	ETOH	3T3[4] Medium	NHK[5] Medium	DMSO	ETOH
Atropine sulfate	200	200	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Boric aid	40	40	200	< 200	Medium	20	20	NT	NT	20	20	NT	NT	2	2	NT	NT
Busulfan	< 2	< 2	40	< 200	DMSO	< 2	< 2	200	NT	< 2	< 2	50[6]	< 200	< 0.2	< 0.2	20	< 200
Caffeine	10	10	20	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Carbamazepine	< 2	< 2	40	< 200	DMSO	0.2	0.2	20	20	< 2	< 2	200	NT	< 0.2	< 0.2	2	< 20
Carbon tetrachloride	2	10	NT	NT	DMSO	20	20	NT	NT	< 0.2	< 0.2	2	NT	20	20	NT	NT
Chloral hydrate	400	400	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Citric acid	400	400	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Cupric sulfate pentahydrate	1	0.5	< 2	2	Medium	2	0.2	< 200	< 200	2	2	NT	NT	0.2	0.2	< 200	NT
Cycloheximide	20	20	400	< 200	Medium	20	20	NT	NT	20	20	NT	NT	2	2	NT	NT
Dibutyl phthalate	< 2	< 2	400	400	DMSO	< 2	< 2	200	NT	< 2	< 2	200	NT	< 2	< 2	200	NT
Dichlorvos	10	10	NT	NT	DMSO	2	2	NT	NT	< 2	< 2	200	NT	2	2	NT	NT
Diethyl phthalate	< 2	< 2	400	400	DMSO	< 2	< 2	200	NT	0.2	< 0.2	200	NT	< 2	< 2	200	NT
Digoxin	0.05	0.05	200	< 200	DMSO	< 2	< 2	200	NT	< 0.2	< 0.2	200	NT	< 2	< 2	200	NT
Dimethylformamide	400	400	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Diquat dibromide monohydrate	200	200	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Disulfoton	< 2	< 2	500	NT	DMSO	< 2	< 2	200	NT	< 2	< 2	200	NT	< 2	< 2	200	NT
Endosulfan	< 0.05	< 0.05	40	NT	DMSO	< 0.2	< 0.2	20	< 200	< 0.2	< 0.2	2	< 200	< 0.2	< 0.2	20	< 200
Epinephrine bitartrate	400	400	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	2	2	NT	NT
Ethanol	200	200	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Fenpropathrin	< 20	< 20	500	NT	DMSO	< 2	< 2	200	NT	< 0.2	< 0.2	200	NT	< 2	< 2	200	NT

Table 5-7   Solubility Results (data presented in mg/mL)

Reference Substance	BioReliance[1]				SMT[2] Selection	ECBC[3]				FAL[3]				IIVS[3]			
	3T3[4] Medium	NHK[5] Medium	DMSO	ETOH		3T3[4] Medium	NHK[5] Medium	DMSO	ETOH	3T3[4] Medium	NHK[5] Medium	DMSO	ETOH	3T3[4] Medium	NHK[5] Medium	DMSO	ETOH
Gibberellic acid	10	10	NT	NT	Medium	2	2	NT	NT	2	2	NT	NT	2	2	NT	NT
Glutethimide	< 2	< 2	500	NT	DMSO	< 2	< 2	200	NT	< 2	< 2	200	NT	< 2	< 2	200	NT
Glycerol	400	400	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Haloperidol	< 20	< 20	40	NT	DMSO	< 0.2	< 0.2	20	< 20	< 0.2	< 0.2	20	< 20	< 2	< 2	20	< 20
Hexachlorophene	0.05	< 0.05	400	400	DMSO	< 2	< 2	200	NT	< 2	< 2	200	NT	< 2	< 2	200	NT
Lactic acid	200	200	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Lindane	< 0.05	< 0.05	400	< 200	DMSO	< 2	< 2	200	NT	< 2	< 2	200	NT	< 0.2	< 0.2	20	< 200
Meprobamate	1	1	200	NT	DMSO	2	2	200	NT	2	2	200	NT	< 0.2	< 0.2	200	NT
Mercury II chloride	0.125	0.125	400	< 200	DMSO	< 2	< 2	200	NT	< 2	< 2	200	NT	< 0.2	< 0.2	200	NT
Methanol	40	40	400	400	DMSO	20	20	NT	NT	20	20	NT	NT	< 2	< 2	200	NT
Nicotine	400	400	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Paraquat	400	400	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Parathion	0.05	< 0.05	400	400	DMSO	< 2	< 2	200	NT	< 2	< 2	200	NT	< 2	< 2	200	NT
Phenobarbital	2	2	200	< 200	DMSO	2	2	NT	NT	< 2	< 2	200	NT	< 2	< 2	200	NT
Phenol	40	40	400	400	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Phenylthiourea	2	2	400	< 200	DMSO	2	< 2	200	NT	20	20	NT	NT	< 2	< 2	200	NT
Physostigmine	2	2	400	200	DMSO	2	2	NT	NT	< 2	< 2	200	NT	< 2	< 2	200	NT
Potassium cyanide	400	400	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Procainamide HCl	200	200	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Propylparaben	0.25	0.25	400	400	DMSO	< 2	< 2	200	NT	< 2	< 2	200	NT	< 2	< 2	200	NT
Sodium arsenite	400	400	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT

Table 5-7 Solubility Results (data presented in mg/mL)

Reference Substance	BioReliance[1]				SMT[2] Selection	ECBC[3]				FAL[3]				IIVS[3]			
	3T3[4] Medium	NHK[5] Medium	DMSO	ETOH		3T3[4] Medium	NHK[5] Medium	DMSO	ETOH	3T3[4] Medium	NHK[5] Medium	DMSO	ETOH	3T3[4] Medium	NHK[5] Medium	DMSO	ETOH
Sodium chloride	200	200	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Sodium dichromate dihydrate	400	400	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Sodium hypochlorite	200	200	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Sodium oxalate	< 0.05	20	0.125	< 0.05	Medium	< 0.2	20	0.2	< 2	20	20	NT	NT	*< 0.2*	*< 0.2*	*< 0.2*	*< 0.2*
Strychnine	< 2	< 2	2	2	Medium	0.2	< 0.2	2	2	0.2	0.2	< 200	< 200	*< 0.2*	*< 0.2*	*< 0.2*	*< 0.2*
Thallium I sulfate	1	0.5	< 2	< 2	Medium	0.2	0.2	< 200	< 200	*< 0.2*	*< 0.2*	*< 0.2*	*< 0.2*	0.2	0.2	< 20	< 200
Trichloroacetic acid	200	200	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
1,1,1-Trichloroethane	10	10	400	400	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Triethylenemelamine	< 2	< 2	2	< 20	DMSO	0.2	0.2	< 200	< 200	< 0.2	< 0.2	2	< 2	*< 0.2*	*< 0.2*	*< 0.2*	*< 0.2*
Triphenyltin hydroxide	< 0.05	< 0.05	10	< 20	DMSO	< 0.2	< 0.2	2	< 20	< 0.2	< 0.2	2	< 200	< 2	< 2	2	< 20
Valproic acid	10	2	NT	NT	DMSO	2	2	NT	NT	< 2	< 2	200	NT	2	< 2	200	NT
Verapamil HCl	< 0.05	0.25	200	NT	DMSO	< 2	< 2	200	NT	< 2	< 2	200	NT	< 0.2	< 0.2	20	NT
Xylene	1	1	500	NT	DMSO	< 2	< 2	200	NT	2	< 2	200	NT	< 2	< 2	200	NT

Table sorted by study phase and alphabetical by reference chemical
[1]Used a different solubility protocol from the *in vitro* cytotoxicity laboratories.
[2]Solvents selected by the SMT for cytotoxicity testing. BioReliance results were used to determine solvents for Phases I and II. Results from all laboratories were used to determine solvents for Phase III. Media were treated as one result. If insoluble in one medium and soluble in DMSO, DMSO was selected.
[3]Used protocol in **Figure 2-7**.
[4]Dulbecco's Modification of Eagle's Medium.
[5]Keratinocyte Growth Medium (KGM® from CAMBREX Clonetics®).
[6]Protocol deviation.
[   ] *In vitro* laboratories agreed on solvent. [   ] *In vitro* laboratories did not agree on solvent. **bold** Protocol did not provide enough information to select a solvent.
NT– not tested.

5.8.2   <u>Solubility Effects on the *In Vitro* NRU Cytotoxicity Test Method Data</u>

The laboratories reported solubility results for the stock solutions for each 3T3 and NHK NRU test. Prior to the additions of the NR dye medium for the NRU test method, the laboratories visually observed the test cultures and documented noticeable precipitate found in the test plates. **Table 5-8** illustrates the existence of solubility issues (in both 3T3 and NHK NRU experiments) as evidenced by the observation of precipitates with some reference substances. Volatility difficulties, indicated by the use of film plate sealers during substance incubation, are also indicated in this table. **Section 3.5** provides additional information on the solubility of specific reference substances.

Table 5-8   Reference Substances with Precipitate (PPT) and Volatility Issues[1]

Reference Substances	3T3 NRU Test Method				NHK NRU Test Method			
	PPT 2X Stock Dilutions	PPT 1X Plate Dilutions	PPT Stock and Plate Dilutions	Volatility	PPT 2X Stock Dilutions	PPT 1X Plate Dilutions	PPT Stock and Plate Dilutions	Volatility
Acetonitrile				X				X
Aminopterin		X			X			
5-Aminosalicylic acid	X							
Arsenic III trioxide	X				X			
Cadmium II chloride		X					X	
Carbamazepine			X					
Carbon tetrachloride			X		X			
Citric acid						X		
Cupric sulfate pentahydrate						X		
Dibutyl phthalate		X					X	
Dichlorvos			X					X
Diethyl phthalate	X						X	
Digoxin			X					
Dimethylformamide						X		
Disulfoton			X				X	
Endosulfan	X			X				X
Ethanol			X					X
Fenpropathrin			X				X	
Gibberellic acid	X				X			
Glutethimide					X			
Lindane			X	X			X	
Lithium I carbonate	X				X			
Nicotine				X				X
Parathion	X					X		
Phenol				X				X
Potassium I chloride		X						
Potassium cyanide		X		X				X
2-Propanol				X				X
Sodium arsenite		X						X
Sodium chloride					X			
Sodium I fluoride		X			X			
Sodium hypochlorite				X				
Sodium oxalate			X			X		
Strychnine	X				X			
Trichloroacetic acid						X		
1,1,1-Trichloroethane	X						X	
Valproic acid	X							
Verapamil HCl					X			
Xylene	X				X			

Table sorted alphabetical by reference substance

[1]Results are based on at least one laboratory having precipitate and volatility issues with a substance. Volatility was denoted by the use of plate sealers during testing. 2X stock dilutions are prepared for each of 8 test substance concentrations. 1X plate dilutions are the result of diluting the 2X stock solutions with medium in the 96-well plate.

5.9   Summary

- Modifications and revisions made to the protocols during Phases I and II contributed to the optimization of the final protocols used in Phase III of the study. The changes did not have a negative impact on the 3T3 and NHK NRU test method data. Generally, changes enhanced the performance of the *in vitro* NRU cytotoxicity test methods and allowed more tests to meet acceptance criteria.

- FAL improved testing quality by modifying the methods used to propagate the NHK cells prior to Phase II testing. Positive control IC_{50} data in Phases II and III from FAL more closely resembled the data from the other laboratories after test methods were optimized.

- Summary test data are presented in tabular and graphical formats. Comparisons of 3T3 IC_{50} values to NHK IC_{50} values show that most values (92%) are within one order of magnitude of each other. Digoxin and aminopterin data had a difference of five orders of magnitude when IC_{50} values are compared between the cell types.

- The BioReliance, ECBC, and IIVS laboratories performed the 3T3 and NHK NRU experiments in compliance with GLP guidelines and submitted quality data. The reference substance quality was maintained throughout the study and lot-to-lot consistency was not a factor in testing.

- Each laboratory followed the same solubility protocol when making reference substance dilutions yet differences in results were present. Judgment of solubility is subjective (as per this protocol).

[This Page Intentionally Left Blank]

36
37
38
39
40
41
42
43
44
45
46
47
48
49
50   ***[This Page Intentionally Left Blank]***
51

51   **6.0   ACCURACY OF THE 3T3 AND NHK NRU TEST METHODS**

52

53   This section discusses the accuracy of the 3T3 and NHK NRU test methods for predicting

54   acute oral systemic toxicity. Accuracy, the agreement between a test method result and an

55   accepted reference value, is a critical component of the ICCVAM evaluation of the validation

56   status of a test method (ICCVAM 2003). Although the 3T3 and NHK NRU test methods are

57   not suitable as replacements for acute systemic toxicity assays, the ability of these assays to

58   correctly predict LD_{50} values are used to evaluate their accuracy. The rationale for

59   evaluating the accuracy of LD_{50} predictions is that the animal savings produced by using

60   these *in vitro* test methods to predict starting doses for acute systemic toxicity assays will be

61   greatest when the starting dose is as close as possible to the "true" LD_{50} value (see **Section**

62   **10** for the evaluation of animal savings).

63

64   The ability of the 3T3 and NHK NRU test methods to correctly predict rodent acute oral

65   systemic toxicity is based on the validity of the *in vitro – in vivo* regression model. It is the

66   *in vivo – in vitro* regression that establishes the relationship between the 3T3 and NHK NRU

67   IC_{50} values and the predicted LD_{50} values that are to be used to set the starting doses for the

68   acute oral systemic toxicity assays in this study.

69

70   Upon review of these regressions, it became apparent that the regression model could be

71   improved. This section discusses the evolution of these improvements. Initially, since the

72   regressions generated by the three laboratories were not statistically different, the data were

73   combined (using a geometric mean IC_{50} of the three individual laboratory geometric mean

74   IC_{50} values) into a single regression for each test method (3T3 and NHK). These regressions,

75   in millimole units, were then compared to the RC millimole regression that was created using

76   rat and mouse oral LD_{50} values from RTECS[®] and IC_{50} values from *in vitro* cytotoxicity

77   assays using multiple cell lines and cytotoxicity endpoints for 347 substances with known

78   molecular weights (Halle 1998). Because the 3T3 and NHK NRU test method regressions

79   were not statistically different from the RC regression, the RC regression was chosen to

80   predict the LD_{50} values from the NRU generated IC_{50} values because it is based on a much

81   larger database.

82   The next steps taken were to improve upon the RC millimole regression's ability to

83   accurately predict LD_{50} values from IC_{50} values, and to make the approach relevant to the

84   testing of mixtures and substances without a known molecular weight in rats, the preferred

85   species for acute oral toxicity testing (EPA 2002b; OECD 2001a; OECD 2001d). To achieve

86   this goal, three new regressions are presented.

87

88   The first regression -- a RC rat-only millimole regression -- utilizes only the 282 substances

89   in the RC dataset of 347 substances that had a reported rat LD_{50} value. The next step was to

90   transform this RC rat-only millimole regression to one based on a weight basis (mg/kg body

91   weight for LD_{50} and µg/mL for IC_{50}) in order to make the regression more generally

92   applicable to the testing of mixtures and substances without a known molecular weight.

93

94   Upon review of this rat-only weight regression, it became apparent that many of the

95   substances with underpredicted toxicity had mechanisms of toxicity that could not be

96   expected to be detected in the 3T3 and NHK cell lines. These mechanisms included

97   neurotoxic and cardiotoxic mechanisms, interference with energy utilization, and agents that

98   alkylate macromolecules. Therefore, the third improved regression presented is based on an

99   RC dataset of 232 substances that have rat LD_{50} data and that excludes the 50 substances

100   which are reported to induce toxicity via one of the above mentioned mechanisms of action.

101

102   The ability of the 3T3 and NHK NRU IC_{50} data to correctly predict rat acute oral LD_{50}

103   values, based on using the RC millimole regression and two of the modified regressions (RC

104   rat-only weight regression and RC rat-only weight regression excluding substances with

105   specific mechanisms of toxicity), was evaluated by determining the extent to which the

106   appropriate GHS acute oral toxicity category was identified for each reference substance.

107   This approach permits an assessment of accuracy specific to each GHS hazard classification

108   category. The results of these analyses are presented in **Section 6.3**. The discordant

109   reference substances from the predictions of GHS acute oral toxicity category are presented

110   in **Appendix L-2**.

111

112

113 The remainder of **Section 6** discusses physical, chemical, and biological characteristics of

114 substances that may have an impact on the accuracy of the 3T3 and NHK NRU test methods.

115

116 **6.1   Accuracy of the 3T3 and NHK NRU Test Methods for Predicting Acute Oral**

117 **Systemic Toxicity**

118

119 Rodent LD_{50} values are used as the reference values for assessing the ability of the 3T3 and

120 NHK NRU test methods to accurately predict acute oral systemic toxicity. The accuracy of

121 the two *in vitro* cytotoxicity test methods is assessed in two ways: (1) by the goodness of fit

122 of the *in vitro* NRU IC_{50} data to the rodent LD_{50} data in linear regression analyses, and (2) by

123 the concordance (i.e., extent of agreement) between the GHS acute oral toxicity categories

124 (UN 2005) assigned based on rodent LD_{50} data and those predicted using *in vitro* NRU IC_{50}

125 data.

126

127 6.1.1   Linear Regression Analyses for the Prediction of *In Vivo* Rodent LD_{50} Values from

128 *In Vitro* NRU IC_{50} Values

129 As described in **Section 5.3.4**, linear regressions for each test method were calculated using

130 log IC_{50} values (mM) versus the corresponding reference log LD_{50} values (mmol/kg)

131 identified in **Table 4-2**. The slopes for all regressions were statistically significantly

132 different from zero ($p < 0.0001$), which indicates a significant relationship between *in vitro*

133 IC_{50} values and the corresponding rodent LD_{50} values.

134

135 Comparison of the individual laboratory regressions to one another with the goodness of fit

136 F-test described in **Section 5.3.3** (under "Generation of Other Linear Regressions") indicated

137 that the laboratory-specific regressions for either *in vitro* NRU cytotoxicity test method were

138 not significantly different from one another (see **Section 7.0** for a more detailed discussion of

139 the results of this analysis). Because the individual laboratory regressions were not

140 significantly different, data were combined into a single regression for each test method

141 using the geometric mean of the mean IC_{50} values determined by each laboratory for each

142 substance (see the "Combined-laboratory" regressions in **Table 6-1** and **Figure 6-1**). The

143   combined-laboratory 3T3 regression yielded a better fit to the reference LD_{50} data (adjusted

144   $R^2 = 0.524$) than the combined-laboratory NHK regression (adjusted $R^2 = 0.455$).

145

146   **Table 6-1   Linear Regression Analyses of the 3T3 and NHK NRU and *In Vivo***

147   **Rodent LD_{50} Test Results[a]**

Laboratory	N^b	Slope	Intercept	Adjusted R^2
3T3 NRU Test Method				
ECBC[c]	69	0.580	0.467	0.531
FAL[c]	67	0.543	0.287	0.432
IIVS[c]	69	0.585	0.467	0.534
Combined-laboratory[d]	70	0.589	0.425	0.524
NHK NRU Test Method				
ECBC[c]	69	0.507	0.405	0.446
FAL[c]	69	0.466	0.427	0.411
IIVS[c]	70	0.513	0.439	0.454
Combined-laboratory[d]	71	0.510	0.452	0.455

148   [a]Log IC_{50} in mM; log LD_{50} in mmol/kg.
149   [b]Number of substances used to calculate regression.
150   [c]Regression based on a single point per substance (i.e., the geometric mean of the
151   within laboratory replicate IC_{50} values and the reference rodent oral LD_{50} from
152   **Table 4-2**).
153   [d]Regression based on a single point per substance (i.e., the geometric mean of the
154   geometric mean IC_{50} values obtained for each laboratory and the reference rodent
155   oral LD_{50} from **Table 4-2**). Data for 70 substances in the 3T3 assay and 71
156   substances in the NHK assay. No laboratory achieved sufficient toxicity for
157   calculation of an IC_{50} for carbon tetrachloride or methanol in the 3T3 NRU test
158   method or for carbon tetrachloride in the NHK NRU test method.
159   Abbreviations: ECBC – US Army Edgewood Chemical Biological Center; FAL –
160   FRAME Alternatives Laboratory; IIVS – Institute for *In Vitro* Sciences

161

162   6.1.2   Comparison of the Combined-Laboratory 3T3 and NHK NRU Regressions to the

163   RC Millimole Regression

164   The NICEATM/ECVAM validation study tested 58 RC substances (see **Figure 3-1**). A

165   comparison of the regression developed for the 3T3 and NHK NRU test results to the RC

166   millimole regression was made to test the assumption of the *Guidance Document* that the RC

167   millimole regression can be obtained with a basal cytotoxicity test method using a single cell

168   type and cytotoxicity endpoint (ICCVAM 2001b). The regression for the 58 substances

169   calculated using the RC IC_{50} and LD_{50} data points is shown in **Figure 6-2**. A graphical

170   comparison of the RC millimole regression for the 58 substances to the 3T3 and NHK

171   combined-laboratory regressions is shown in **Figure 6-3**. A statistical comparison of slope

172

172   **Figure 6-1   Combined-Laboratory 3T3 and NHK Regressions**

3T3

log LD50 = 0.589 log IC50 + 0.425
adjusted R^2 = 0.524

NHK

log LD50 = 0.510 log IC50 + 0.452
adjusted R^2 = 0.455
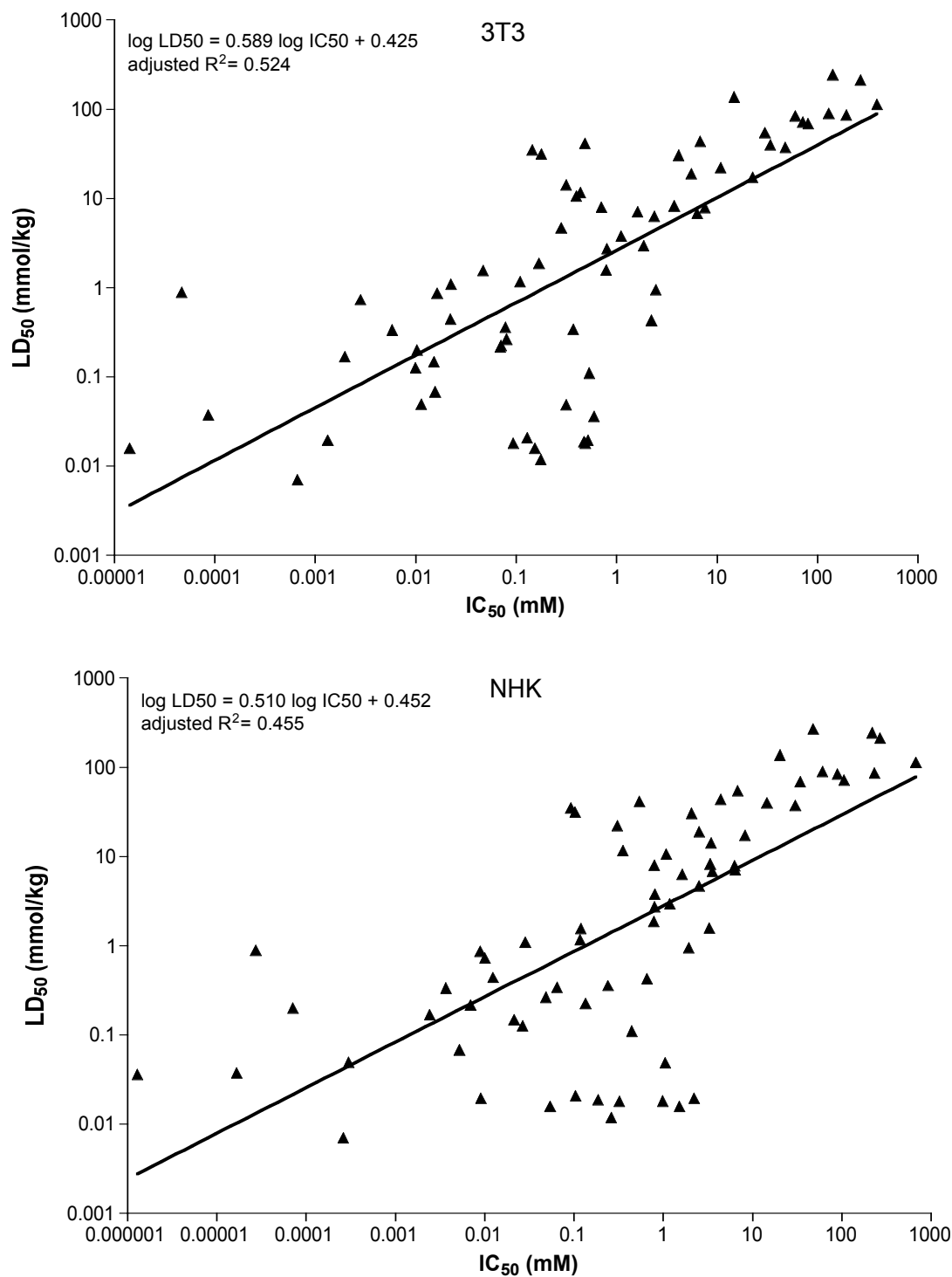
173
174   Solid lines show the combined-laboratory regressions for each test method (see **Table 6-1**).
175

175   and intercept (simultaneously) using an F test showed that neither the 3T3 regression (p =

176   0.929) nor the NHK regression (p = 0.144) was different from the 58 RC substance

177   regression.

178

179   **Figure 6-2   Regression for the 58 RC Substances Using RC Data**

log LD50 = 0.659 log IC50 + 0.322
adjusted R^2 = 0.683

(Axis labels: Y-axis LD_{50} (mmol/kg); X-axis IC_{50} (mM))

180
181
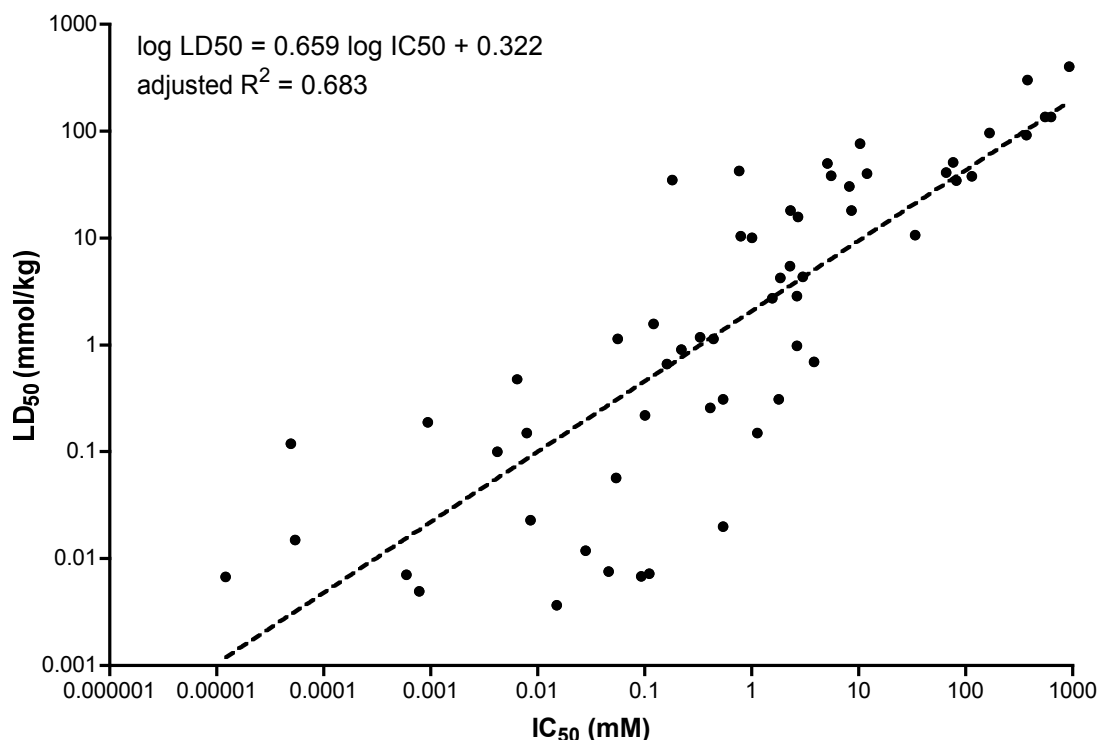
182   **6.2   Improving the Prediction of *In Vivo* Rodent LD$_{50}$ Values from *In Vitro* NRU**

183   **IC$_{50}$ Data**

184

185   Since the RC and the 3T3 and NHK NRU IC_{50} – rodent acute oral LD_{50} regressions were not

186   significantly different, the next step was an attempt to improve upon the RC millimole

187   regression for the prediction of LD_{50} from IC_{50}. The RC data were used to develop three new

188   regressions. For reference, the original RC millimole regression, log LD_{50} (mmol/kg) =

189   0.435 x log IC_{50x} (mM) + 0.625 (Halle 1998), is shown in **Table 6-2** and **Figure 6-4**.

190

190   **Figure 6-3   Regression for the 58 RC Substances with the 3T3 and NHK NRU**

191   **Regressions**

192
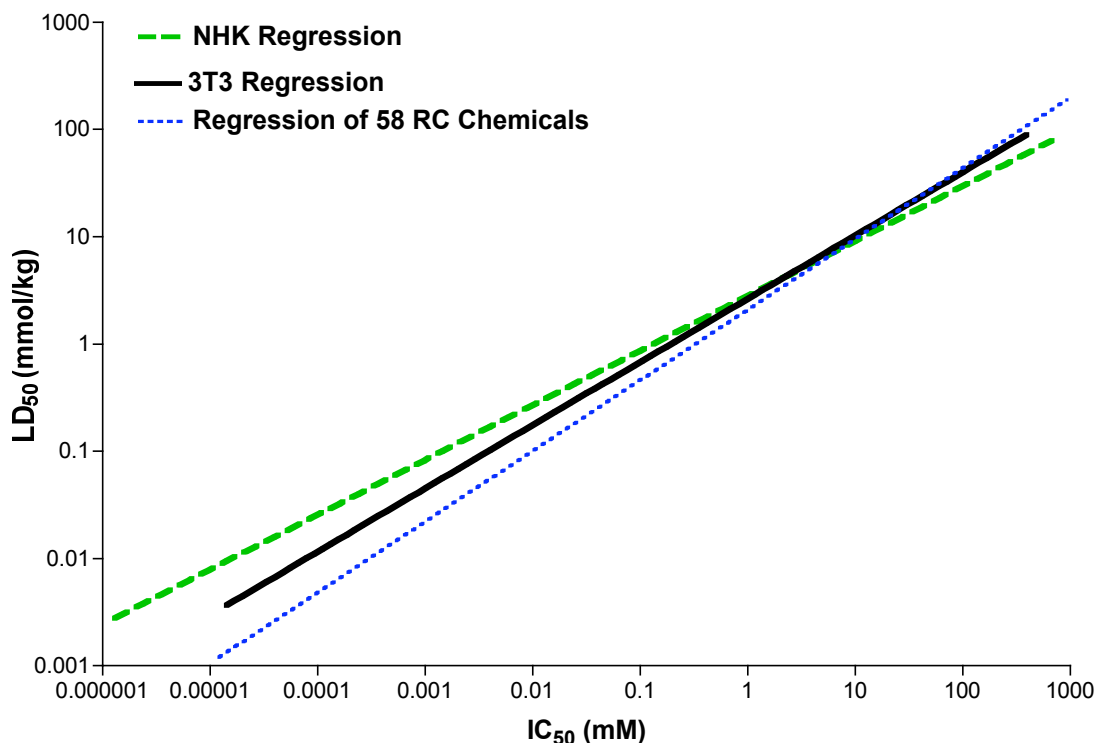
193   Regression for the 58 RC substances using RC data is log LD_{50} = 0.659 log IC_{50} + 0.323 (adjusted R^2
194   = 0.683). The combined-laboratory 3T3 NRU regression, which uses data for 70 substances, is log
195   LD_{50} = 0.589 log IC_{50} + 0.425 (adjusted R^2 = 0.524) (from **Table 6-1**). The combined-laboratory
196   NHK NRU regression, which uses data for 71 substances, is log LD_{50} = 0.510 log IC_{50} + 0.452
197   (adjusted R^2 = 0.455) (from **Table 6-1**). No laboratory achieved sufficient toxicity for calculation of
198   an IC_{50} for carbon tetrachloride or methanol in the 3T3 NRU test method or for carbon tetrachloride in
199   the NHK NRU test method.

200

201   6.2.1   <u>The RC Rat-Only Regression in Millimolar Units</u>

202   The first regression used the RC data only for the 282 substances with rat LD_{50} data (i.e., the

203   regression excluded the substances with mouse LD_{50} data) using the original units of mM for

204   IC_{50} and mmol/kg for LD_{50} (see **Table 6-2** and **Figure 6-4**). Rat data only were used

205   because:

206   •   rats and mice may not have the same sensitivity to individual substances,

207   regardless of the high correlation of a subset of 173 RC substances with both rat

208   and mouse LD_{50} data (r_s = 0.88; p < 0.0001) (see **Section 4.1.4**)

209   • the majority of LD_{50} data used in the RC millimole regression were from studies

210   using rats (282 rat data points and 65 mouse data points) (Halle 1998)

211   • the great majority of acute oral systemic toxicity testing is performed with rats

212

213 **Table 6-2**   **Linear Regression Analyses to Improve the Prediction of Rodent LD_{50}**

214   **from *In Vitro* NRU IC_{50} Using the RC Regression[a]**

Data Used	Slope	Intercept	Adjusted R^2
347 RC substances with rat and mouse LD_{50} data – millimole units[c]	0.435	0.625	0.450[d]
282 RC substances with rat LD_{50} data – millimole units[c]	0.439	0.621	0.451
282 RC substances with rat LD_{50} data – weight units[e]	0.372	2.024	0.322
232 RC substances with rat LD_{50} data (excluded 50 substances with specific mechanisms of action[f]) – weight units[e]	0.357	2.194	0.353

215 [a]Slopes of all regressions were significantly different ($p < 0.05$) from zero at $p < 0.0001$.
216 [b]Simultaneous comparison of slopes and intercepts using an F test. Significance denoted by $p < 0.05$.
217 [c]IC_{50} in mM; LD_{50} in mmol/kg.
218 [d]Calculated from RC data (i.e., not reported by Halle [1998]).
219 [e]IC_{50} in µg/mL; LD_{50} in mg/kg.
220 [f]See the text for the applicable mechanisms and **Appendix K-3** for the identified substances.
221

222 **Table 6-2** shows that the regression using rat LD_{50} data only was almost identical to the

223 original RC millimole regression which used both rat and mouse LD_{50} data. The slope

224 changed from 0.435 for the RC millimole regression to 0.439 and the intercept changed from

225 0.625 to 0.621.

226

227 6.2.2   The RC Rat-Only Regression in Weight Units

228 The second regression used the same RC data for the 282 substances with rat LD_{50} data, but

229 was calculated with weight units rather than millimolar units (see **Table 6-2** and **Figure 6-5**).

230 Weight units (i.e., mg/kg for the LD_{50} and µg/mL for the IC_{50}) were selected for the units of

231 measurement because

232   • millimole units are not applicable to mixtures and unknown substances

233   • they are most practical [i.e., in all regulatory systems, hazard classification is

234   based on LD_{50} values expressed in mg/kg (see **Table 1-2**)]

235

235   **Figure 6-4   RC Regression (a) and RC Rat Regression (b) Using Millimole Units**

a.

log LD50 = 0.435 log IC50 + 0.625
adjusted R^2= 0.450

b.

log LD50 = 0.439 log IC50 + 0.621
adjusted R^2= 0.451

236
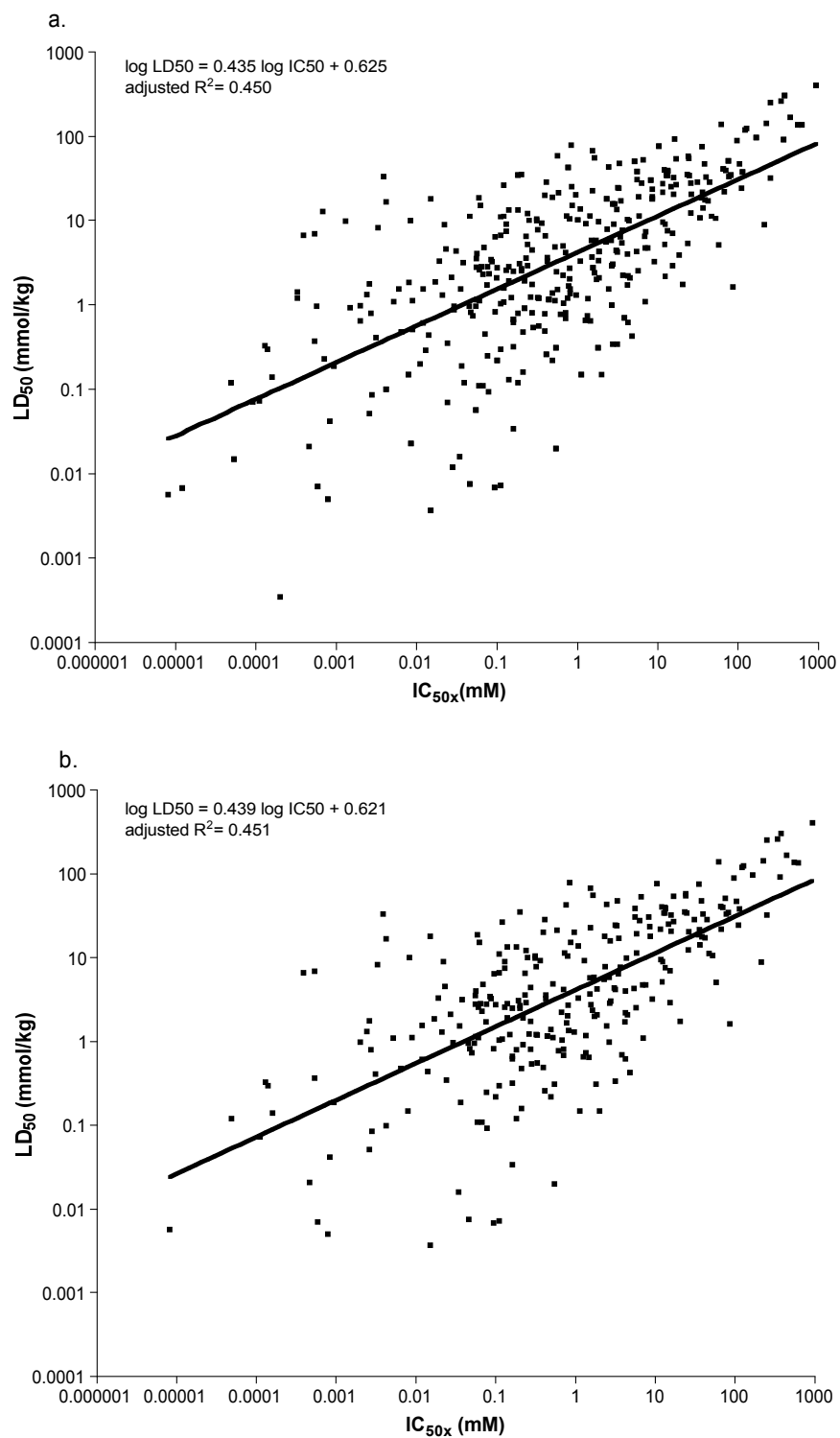
237   6.2.3   The RC Rat-Only Regression in Weight Units Excluding Substances with Specific
238   Mechanisms of Toxicity

239   The third regression was a further refinement on the weight-unit regression developed from

240   the 282 RC substances with rat LD_{50} data. It excluded the RC substances for which the

241   mechanisms of toxic action were not expected to be active in the 3T3 and NHK cell cultures.

242   This reduced the number of data points from 282 to 232 RC substances for the calculation of

243   the regression (see **Table 6-2** and **Figure 6-5**). The third regression was significantly

244   different ($p < 0.05$) from the RC rat-only weight regression when slopes and intercepts were

245   simultaneously compared (F test; $p = 0.0063$). The idea for the further refinement for the rat

246   RC millimole regression came from the evaluation of discordant substances (i.e., those

247   greater than 0.699 or 0.5 log from the regression) when the 3T3 and NHK NRU data were

248   used with the RC millimole regression (see **Appendix L-1**). For the 3T3 NRU, 13/30 (43%)

249   of the discordant substances had mechanisms of toxicity that were not expected to be active

250   in the 3T3 cell cultures. For the NHK NRU, 13/31 (42%) of the discordant substances had

251   mechanisms of toxicity that were not expected to be active in the NHK cell cultures.

252

253   *Development of the RC Rat-Only Weight Regression Excluding Substances with Specific*

254   *Mechanisms of Toxicity*

255   Mechanism of action data for the 282 RC substances with rat LD_{50} values were obtained

256   from *Casarett & Doull's Toxicology* (Casarett et al. 2001) and the following Internet sources:

257   HSDB (NLM 2005); Haz-Map (NLM 2005); Pesticide Action Network [PAN] Pesticides

258   Database (PAN North America 2005); and IPCS INTOX Database (Canadian Centre for

259   Occupational Health and Safety 2005) (see **Appendix K-3**). Mechanism of action

260   information could not be found for all substances. For 35 of the 282 (12%) substances, only

261   the product class could be identified; for seven (3%) substances, no information was found.

262   Examination of the RC rat database revealed the following.

263   • Of the 282 substances, 73 (26%) were outliers[1] (i.e., log observed – log predicted

264        $LD_{50} > 0.699$ as defined for the RC millimole regression).

265

---

[1]Substance "outliers" are often referred to as discordant chemicals. Substance outliers are different from the
replicate "outliers" described in **Sections 5.2** and **5.3**, which were extreme values in a set of replicate data. See
Section 13 for definitions.

265   **Figure 6-5**   **RC Rat-Only Regression (a) and RC Rat-Only Regression after**

266   **Excluding 50 Substances with Specific Mechanisms of Toxicity (b) Using**

267   **Weight Units**

a.

$\log LD50 = 0.372 \log IC50 + 2.024$
adjusted $R^2 = 0.322$

b.

$\log LD50 = 0.357 \log IC50 + 2.194$
adjusted $R^2 = 0.353$

268
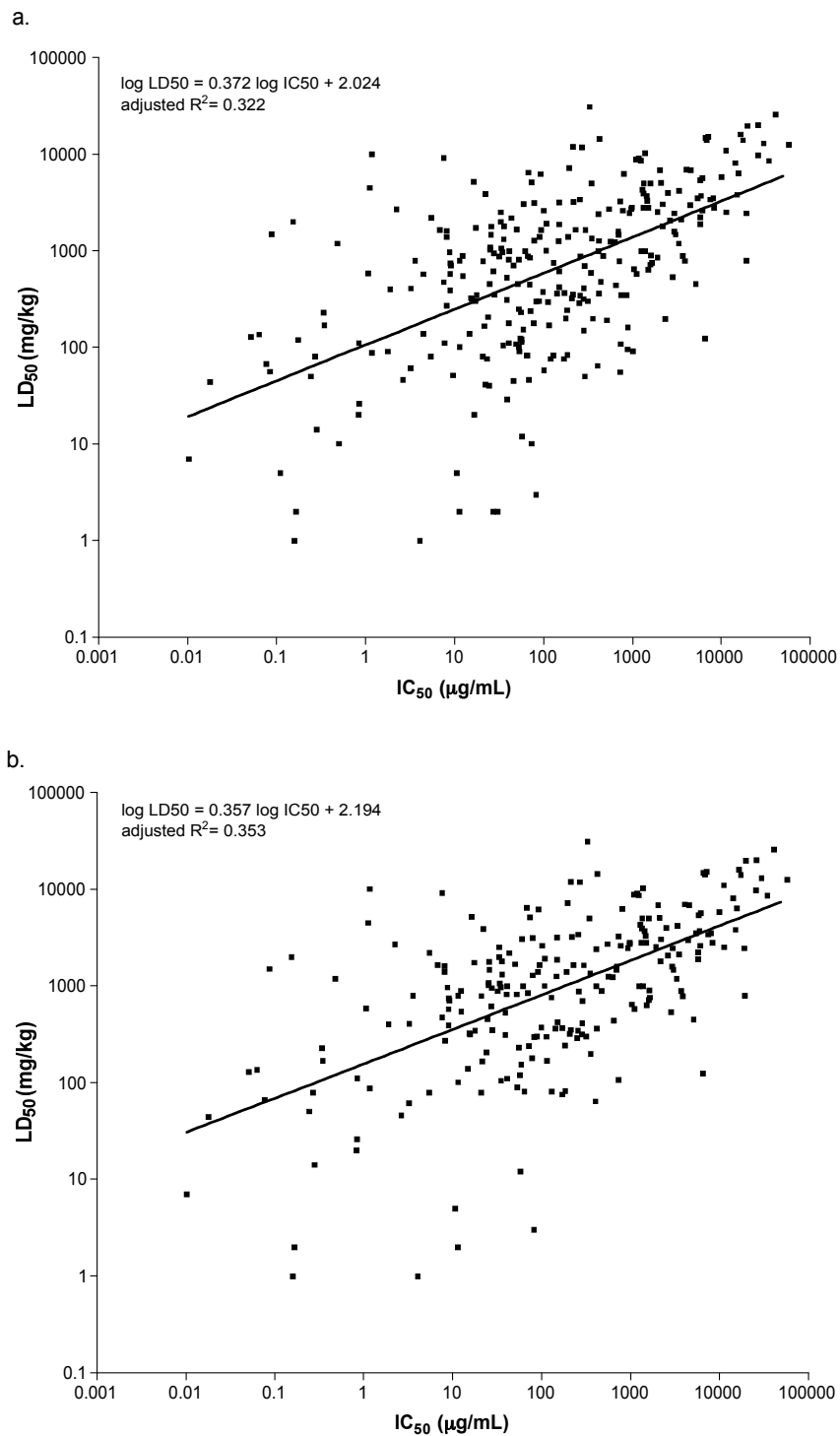
269   • For 40 (55%) of the 73 substances, *in vivo* toxicity was underpredicted; for 33

270   (45%) of the 73 substances, *in vivo* toxicity was overpredicted.

271   • All underpredicted substances were very toxic, with $LD_{50} \leq 200$ mg/kg.

272   • The discordant status of 65% (26/40) of the underpredicted substances could be

273   explained by four general mechanisms

274   o   neurotoxic (i.e., cholinesterase inhibitor, affects CNS nicotinic receptor,

275   or otherwise neurotoxic by a mechanism other than membrane

276   destabilization such as that produced by a solvent)

277   o   interferes with energy utilization (i.e., interferes with ATP synthesis,

278   inhibits ADP phosphorylation, or uncouples oxidative phosphorylation,

279   or is a metabolic poison)

280   o   cardiotoxic via specific mechanisms (i.e., positive inotropic action,

281   calcium channel blocker)

282   o   alkylates cellular proteins and other macromolecules (i.e., covalently

283   binds to enzymes and other proteins to disrupt normal function)

284

285 Substances with such mechanisms would not be expected to exert their toxic mechanisms in

286 the 3T3 and NHK cells and thus, they would be expected to fit the RC millimole regression

287 poorly, as evidenced by their discordant status. A new regression was calculated after the

288 exclusion of all substances in the RC database known to act by these four mechanisms; this

289 included the 26 underpredicted substances and 24 other substances that were not identified as

290 outliers. The substances excluded from the RC rat weight regression are identified in

291 **Appendix K-3**.

292

293 **6.3   Accuracy of the 3T3 and NHK NRU Test Methods for Toxicity Category**

294   **Predictions**

295

296 The 3T3 and NHK NRU test methods are not suitable as replacements for acute oral systemic

297 toxicity assays. However, the use of *in vitro* NRU test methods to reduce animal use for

298 acute oral systemic toxicity assays depends upon their accuracy for the prediction of LD_{50}

299 values. NRU-predicted LD_{50} values were determined by using the *in vitro* NRU IC_{50} values

300   in the regressions presented in **Table 6-2.** The predicted LD_{50} values were then used to

301   assign each substance to a predicted GHS acute oral category (UN 2005). The accuracy of

302   the 3T3 and NHK NRU test methods for predicting GHS toxicity categories was determined

303   by comparison with categorization based on *in vivo* rodent LD_{50} data. This accuracy

304   evaluation approach was used because the regulatory use of acute systemic toxicity test

305   results is for the purpose of hazard classification and labelling of products to protect handlers

306   and consumers.

307

308   The following regressions from **Section 6.2** were evaluated for accuracy of GHS oral toxicity

309   category predictions:

310   • RC millimole regression

311   • RC rat-only weight regression

312   • RC rat-only weight regression excluding substances with specific mechanisms

313   of toxicity

314   The regression calculated using the rat RC data in millimole units (**Section 6.2.1**) was not

315   evaluated separately since it was very similar to the original RC millimole regression, which

316   used both rat and mouse data. As **Table 6-2** shows, the slopes and intercepts varied only in

317   the thousandths digits.

318

319   Data for the same reference substances were evaluated for each regression. Forty-six

320   substances were evaluated for the 3T3 NRU test method and 47 substances were evaluated

321   for the NHK NRU test method. Of the original 72 substances tested, epinephrine bitartrate,

322   colchicine, and propylparaben were excluded because they were removed from the

323   calculation of the RC rat-only weight regression due to the lack of rat oral reference LD_{50}

324   data. The 21 substances with specific mechanisms of toxicity in **Table 6-3** were excluded

325   from all analyses to be consistent with those removed from the RC rat-only weight regression

326   excluding substances with specific mechanisms of toxicity. These substances have known

327   mechanisms of toxicity that are not expected to be active in the 3T3 and NHK cell cultures.

328

328   **Table 6-3   Substances Deleted from the Evaluations of the 3T3 and NHK NRU Test**

329   **Methods and Regressions Due to Mechanisms of Toxicity Not Expected to**

330   **Be Active in the 3T3 and NHK Cell Cultures**

Substance	Mechanism of Toxicity[1]
Neurotoxic	
Amitriptyline HCl	Blocks norepinephrine, 5-hydroxytryptamine, and dopamine presynaptic uptake; prevents reuptake of heart norepinephrine.
Atropine sulfate	Antimuscarinic, anticholinergic action; competitive antagonism of anticholinesterase at cardiac & CNS receptor sites.
Caffeine	Inhibition of phosphodiesterase leading to AMP accumulation, translocation of intracellular Ca^{++}, adenosine receptor antagonism, neurotoxic.
Carbamazepine	Therapeutically decreases firing of noradrenergic neurons.
Chloral hydrate	Potentiation of $GABA_A$ receptor activity, inhibition of N-methyl-D-aspartate activity, & modulation of 5-hydroxytryptamine$_3$ receptor-mediated depolarization of the vagas nerve[2].
Dichlorvos	Inhibition of acetylcholinesterase resulting in acetylcholine accumulation in CNS & effector organs.
Disulfoton	Inhibition of acetylcholinesterase resulting in acetylcholine accumulation in CNS & effector organs.
Endosulfan	Affects brain neurotransmitter levels[3].
Fenpropathrin	Delays closure of sodium channel causing persistent depolarization of membrane.
Glutethimide	CNS depression, anticholinergic activity.
Haloperidol	Blocks dopamine receptors.
Lindane	CNS depression through inhibition of GABA receptor linked chloride channel at the picrotoxin binding site, leading to blockade of chloride influx into neurons.
Nicotine	Cholinergic block causing polarization of CNS and PNS synapses.
Parathion	Inhibition of acetylcholinesterase resulting in acetylcholine accumulation in CNS & effector organs.
Phenobarbital	CNS depression through inhibition of GABA synapses, inhibits hepatic NADH cytochrome oxidoreductase.
Physostigmine	Inhibition of acetylcholinesterase resulting in acetylcholine accumulation in CNS & effector organs.
Strychnine	Increases glutamic acid in the CNS.
Interferes with Energy Utilization	
Paraquat	Multisystem failure due to depletion of superoxide dismutase, with formation of free radicals & lipid peroxidation; lung fibrosis due to accumulation.
Potassium cyanide	General enzyme inhibition, high affinity for Fe^{+++}, inhibits cell respiration by inhibition of cytochrome oxidase.
Cardiotoxic	
Procainamide HCl	Slows impulse conduction in the heart[4].
Verapamil HCl	Inhibition of transmembrane Ca^{++} flux in excitatory tissues, alpha-adrenergic blockade.

331   Abbreviations: NA = not available or information not found; CNS = central nervous system; GABA = gamma
332   aminobutyric acid; PNS = peripheral nervous system; NADH = nicotine adenine dinucleotide (reduced).
333   [1]Ekwall et al. (1998) or Hazardous Substances Data Bank (NLM 2001, 2002) unless otherwise noted.
334   [2]EPA (2000b).
335   [3]ATSDR (2000a).
336   [4]Hardman et al. (1996).

337   Carbon tetrachloride and methanol were excluded from the 3T3 NRU evaluations because no

338   laboratory attained sufficient toxicity in any test for the calculation of an IC_{50}. Carbon

339   tetrachloride was also excluded from the NHK NRU evaluations because no laboratory

340   attained sufficient toxicity in any test for the calculation of an IC_{50}.

341

342   The tables providing accuracy information in this section (**Tables 6-4** to **6-6**) are divided into

343   top and bottom parts that provide accuracy data for the 3T3 and NHK NRU test methods,

344   respectively. For each part, the toxicity categories corresponding to the *in vivo* LD_{50} data are

345   provided in rows that are labeled on the far left side of the table. The toxicity categories

346   predicted by the *in vitro* NRU assays (and associated regressions) are provided in columns

347   that are labeled across the top of each part (i.e., 3T3 or NHK NRU-predicted toxicity

348   category) of the table. The numbers at the intersections of the *in vivo* LD_{50} rows and 3T3 or

349   NHK NRU-predicted toxicity category columns are the numbers of substances with *in vitro*

350   category predictions that correspond to the various *in vivo* categories. The right sides of the

351   tables also provide summaries containing, for each *in vivo* toxicity category and for the total

352   number of substances evaluated: number of substances, the accuracy of the 3T3 or NHK

353   NRU prediction, and the percentage of substances for which toxicity has been overpredicted

354   and underpredicted by the *in vitro* NRU methods. In each of the 3T3 NRU and NHK NRU

355   sections of the table, a summary of predictivity[2] is also provided for each predicted toxicity

356   category along with the percentage of substances with category (i.e., toxicity) underpredicted

357   and overpredicted.

358

359   6.3.1   <u>Prediction of Toxicity Category by the 3T3 and NHK NRU Test Methods Using the</u>

360   <u>RC Millimole Regression</u>

361   **Table 6-4** shows the concordance of the observed (i.e., *in vivo*) and predicted GHS acute oral

362   toxicity categories (UN 2005) for each *in vitro* NRU cytotoxicity test method using the

363   geometric mean IC_{50} values (of the three laboratories) in the RC millimole regression, log

364   LD_{50} (mmol/kg) = 0.435 x log IC_{50} (mM) + 0.625. Accuracy is the agreement of the

---

[2] Proportion of *in vivo* category matches for all substances with *in vitro* predictions for a particular category. Predictivity is an indicator of test accuracy (ICCVAM 2003).

365 **Table 6-4**   **Prediction of GHS Toxicity Category[1] by the 3T3 and NHK NRU Test Methods and the**

366   **RC Millimole Regression[2]**

Reference Rodent LD$_{50}$[3]	3T3 NRU-Predicted Toxicity Category						Total	Accuracy	Toxicity Overpredicted	Toxicity Underpredicted
	< 5	5 – 50	50 – 300	300 – 2000	2000 – 5000	> 5000				
< 5	0	3	1	3	0	0	7[4]	0%	0%	100%
5 –50	0	1	3	1	1	0	6[5]	17%	0%	83%
50 – 300	0	0	4	2	0	0	6[6]	67%	0%	33%
300 – 2000	0	0	0	6	0	0	6[7]	100%	0%	0%
2000 – 5000	0	0	0	11	0	0	11[8]	0%	100%	0%
> 5000	0	0	0	6	3	1	10[9,10]	10%	90%	0%
Total	0	4	8	29	4	1	46	26%	43%	30%
Predictivity	0%	25%	50%	21%	0%	100%				
Category Underpredicted	0%	0%	0%	59%	75%	0%				
Category Overpredicted	0%	75%	50%	21%	25%	0%				
Reference Rodent LD$_{50}$[3]	NHK NRU-Predicted Toxicity Category						Total	Accuracy	Toxicity Overpredicted	Toxicity Underpredicted
	< 5	5 – 50	50 – 300	300 – 2000	2000 – 5000	> 5000				
< 5	0	1	3	2	1	0	7[4]	0%	0%	100%
5 – 50	0	3	3	0	0	0	6[5]	50%	0%	50%
50 – 300	0	1	3	2	0	0	6[6]	50%	17%	33%
300 – 2000	0	0	0	6	0	0	6[7]	100%	0%	0%
2000 – 5000	0	0	0	10	1	0	11[8]	9%	91%	0%
> 5000	0	0	0	6	5	0	11[10]	0%	100%	0%
Total	0	5	9	26	7	0	47	28%	47%	26%
Predictivity	0%	60%	33%	23%	14%	0%				
Category Underpredicted	0%	20%	0%	62%	71%	0%				
Category Overpredicted	0%	20%	67%	15%	14%	0%				

367 [1]GHS-Globally Harmonized System of Classification and Labelling of Chemicals with LD_{50} in mg/kg (UN 2005).
368 < 5: $LD_{50} \leq 5$ mg/kg
369 5 – 50: $5 < LD_{50} \leq 50$ mg/kg
370 50 – 300: $50 < LD_{50} \leq 300$ mg/kg
371 300 – 2000: $300 < LD_{50} \leq 2000$ mg/kg
372 2000 – 5000: $2000 < LD_{50} \leq 5000$ mg/kg
373 > 5000: $LD_{50} > 5000$ mg/kg
374 [2]The RC millimole regression is log LD_{50} (mmol/kg) = log IC_{50} (mM) X 0.435 + 0.625. Numbers in table represent number of
375 substances.
376 [3]Reference oral LD_{50} values from **Table 3-2**.
377 [4]Epinephrine bitartrate excluded because no rat LD_{50} was identified. Disulfoton, parathion, strychnine and physostigmine
378 excluded based on mechanism of toxicity (see **Table 6-3**).
379 [5]Colchine excluded because no rat LD_{50} was identified. Dichlorvos, endosulfan, fenpropathrin, nicotine, and potassium cyanide
380 excluded based on mechanism of toxicity (see **Table 6-3**).
381 [6]Caffeine, haloperidol, lindane, paraquat, phenobarbital, and verapamil HCl excluded based on mechanism of toxicity (see **Table
382 6-3**).
383 [7]Amitriptyline, atropine sulfate, carbamazepine, chloral hydrate, glutethimide, and procainamide HCl excluded based on
384 mechanism of toxicity (see **Table 6-3**).
385 [8]Carbon tetrachloride excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50}.
386 [9]Methanol excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50}.
387 [10]Propylparaben excluded because no rat LD_{50} was identified.
388

389   category predictions with those based on the reference rodent LD_{50} values in **Table 3-2**,

390   which are the values used for the original classification of the test substances. Substances for

391   which the *in vitro* toxicity category prediction does not match the *in vivo* determined toxicity

392   category are considered discordant substances for the GHS toxicity category predictions.

393

394   *In Vitro – In Vivo Concordance Using the RC Millimole Regression*

395   The overall accuracy of the 3T3 NRU test method for correctly predicting GHS toxicity

396   classification category using the RC millimole regression was 26% (12/46 substances)

397   (**Table 6-4**). *In vivo* toxicity was overpredicted for 43% (20) and underpredicted for 30%

398   (14) of the 46 substances. For this analysis, in terms of each GHS toxicity classification

399   category:

400   •   0 (0%) of seven substances with LD_{50} < 5 mg/kg was correctly predicted

401   •   1 (17%) of six substances in the 5 < LD_{50} ≤ 50 mg/kg category was correctly

402   predicted

403   •   4 (67%) of six substances in the 50 < LD_{50} ≤ 300 mg/kg category were correctly

404   predicted

405   •   6 (100%) of six substances in the 300 < LD_{50} ≤ 2000 mg/kg category were

406   correctly predicted; however, this toxicity category was also predicted for 23

407   other substances (79%; 23/29) that did not match this category *in vivo*. Thus,

408   the predictivity for this category was 21% (6/29 substances predicted for this

409   category matched the *in vivo* category).

410   •   0 (0%) of the 11 substances in the 2000 < LD_{50} ≤ 5000 mg/kg category were

411   correctly predicted

412   •   1 (10%) of the 10 substances in the LD_{50} > 5000 mg/kg range was correctly

413   predicted

414

415   The overall accuracy of the NHK NRU cytotoxicity test method for correctly predicting the

416   GHS toxicity classification, when the prediction was based on the RC millimole regression,

417   was 28% (13/47 substances) (see **Table 6-4**). Toxicity was overpredicted for 47% (22) and

418   underpredicted for 26% (12) of the 47 substances. The pattern of concordance between *in*

419   *vitro* and *in vivo* results for the NHK assay with the RC millimole regression was similar to

420   the 3T3 results with the exception that two more substances were correctly in the $5 < LD_{50} \leq$

421   50 mg/kg category. For this analysis, in terms of each GHS toxicity classification category:

422   •   0 (0%) of seven substances with $LD_{50} < 5$ mg/kg were correctly predicted

423   •   3 (50%) of six substances in the $5 < LD_{50} \leq 50$ mg/kg and in the

424        $50 < LD_{50} \leq 300$ mg/kg categories were correctly predicted

425   •   6 (100%) of six substances in the $300 < LD_{50} \leq 2000$ mg/kg category were

426   correctly predicted; however, this toxicity category was also predicted for 20

427   (77%; 20/26) other substances with *in vivo* data that did not match the category.

428   Thus, the predictivity for this category was 23%.

429   •   1 (9%) of 11 substances in the $2000 < LD_{50} \leq 5000$ mg/kg category was

430   correctly predicted

431   •   0 (0%) of 11 substances in the $LD_{50} > 5000$ mg/kg range were correctly

432   predicted.

433

434   For both *in vitro* NRU cytotoxicity test methods, when predicted GHS toxicity categories did

435   not match the reference rodent GHS toxicity category, the RC millimole regression generally

436   underpredicted toxicity for substances in the highest toxicity (i.e., lowest LD_{50}) categories

437   and overpredicted toxicity for substances in the lowest toxicity (i.e., highest LD_{50}) categories

438   (see **Table 6-4**). While substances at the very low and very high ends of the toxicity range

439   were poorly predicted, those in the middle of the toxicity range (i.e., $300 < LD_{50} \leq 2000$

440   mg/kg) were predicted quite well.

441

442   *Discordant Substances for Prediction of Toxicity Category by the 3T3 and NHK NRU Test*

443   *Methods and the RC Millimole Regression*

444   **Appendix L-2** identifies the discordant substances for which the *in vitro* predicted GHS

445   toxicity category did not match the GHS toxicity category assigned based on the reference

446   rodent LD_{50} data in **Table 3-2**. For the 3T3 NRU test method, the toxicity was

447   underpredicted for 14 (30%) and overpredicted for 20 (43%) of the 34 discordant substances.

448   For the NHK NRU test method, toxicity was underpredicted for 12 (35%) and overpredicted

449   for 22 (65%) of the 34 discordant substances. The fact that there were more substances for

450   which toxicity was overpredicted is a result of the removal of substances with specific

451 mechanisms of toxicity that were not expected to be active in the 3T3 and NHK cell cultures.

452 The toxicity for most of these substances would have been underpredicted. **Figure 3-1**

453 shows that most of the RC substances selected for testing in the NICEATM/ECVAM

454 validation study are below the RC millimole regression line. Thus, the RC millimole

455 regression is expected to predict lower toxicity (i.e., a higher LD_{50}) for these substances.

456

457 6.3.2   Prediction of Toxicity Category by the 3T3 and NHK NRU Test Methods Using the

458   RC Rat-Only Weight Regression

459 **Table 6-5** shows the concordance of the observed and predicted GHS acute oral toxicity

460 categories for each *in vitro* NRU test method using the geometric mean IC_{50} values (of the

461 three laboratories) and the RC rat only weight regression from **Table 6-2**. The regression

462 formula for the RC rat-only weight regression was log LD_{50} (mg/kg) = log IC_{50} (μg/mL) x

463 0.372 + 2.024. Accuracy is the agreement of the *in vitro* NRU cytotoxicity GHS toxicity

464 category predictions with those based on the reference rat oral LD_{50} values from **Table 4-2**.

465

466 *In Vitro – In Vivo Concordance with the RC Rat-Only Weight Regression*

467 The overall accuracy of the 3T3 NRU test method with the RC rat-only weight regression

468 was 35% (16) for the results from 46 substances (**Table 6-5**). The toxicity was overpredicted

469 for 41% (19) and underpredicted for 24% (11) of the 46 substances. For this analysis, in

470 terms of each GHS toxicity classification category:

471   •   0 (0%) of four substances with $LD_{50} < 5$ mg/kg were correctly predicted

472   •   1 (14%) of seven substances in the $5 < LD_{50} \le 50$ mg/kg GHS toxicity category

473   was correctly predicted

474   •   4 (80%) of five substances in the $50 < LD_{50} \le 300$ mg/kg GHS toxicity category

475   were correctly predicted; however, since seven other substances were also

476   predicted for this category, the predictivity was 36% (4/11)

477   •   7 (78%) of nine substances in the $300 < LD_{50} \le 2000$ mg/kg GHS toxicity

478   category were predicted correctly. Since a total of 22 substances were predicted

479   for this category, the predictivity was 32% (7/22).

480

481 **Table 6-5   Prediction of GHS Toxicity Category[1] by the RC Rat-Only Weight Regression[2]**

Reference Rodent LD_{50}[3]	3T3 NRU-Predicted Toxicity Category						Total	Accuracy	Toxicity Overpredicted	Toxicity Underpredicted
	< 5	5 – 50	50 – 300	300-2000	2000-5000	> 5000				
< 5	0	0	2	2	0	0	4[4]	0%	0%	100%
5 – 50	0	1	4	2	0	0	7[5]	14%	0%	86%
50 – 300	0	0	4	1	0	0	5[6]	80%	0%	20%
300 – 2000	0	1	1	7	0	0	9[7]	78%	22%	0%
2000 – 5000	0	0	0	5	4	0	9[8]	44%	56%	0%
> 5000	0	0	0	5	7	0	12[9,10]	0%	100%	0%
Total	0	2	11	22	11	0	46	35%	41%	24%
Predictivity	0%	50%	36%	32%	36%	0%				
Category Underpredicted	0%	50%	9%	46%	64%	0%				
Category Overpredicted	0%	0%	55%	23%	0%	0%				

Reference Rodent LD_{50}[3]	NHK NRU-Predicted Toxicity Category						Total	Accuracy	Toxicity Overpredicted	Toxicity Underpredicted
	< 5	5 – 50	50 – 300	300 – 2000	2000 – 5000	> 5000				
< 5	0	1	2	1	0	0	4[4]	0%	0%	100%
5 – 50	0	1	4	2	0	0	7[5]	14%	0%	86%
50 – 300	0	1	3	1	0	0	5[6]	60%	20%	20%
300 – 2000	0	1	0	8	0	0	9[7]	89%	11%	0%
2000 – 5000	0	0	0	8	1	0	9[8]	11%	89%	0%
> 5000	0	0	0	6	6	1	13[10]	8%	92%	0%
Total	0	4	9	26	7	1	47	30%	47%	23%
Predictivity	0%	25%	33%	31%	14%	100%				
Category Underpredicted	0%	50%	0%	54%	86%	0%				
Category Overpredicted	0%	25%	67%	15%	0%	0%				

482 [1]Globally Harmonized System of Classification and Labelling of Chemicals with LD_{50} in mg/kg (UN 2005).
483 < 5: $LD_{50} \leq 5$ mg/kg
484 5 – 50: $5 < LD_{50} \leq 50$ mg/kg
485 50 – 300: $50 < LD_{50} \leq 300$ mg/kg

486   300 – 2000: $300 < LD_{50} \leq 2000$ mg/kg
487   2000 – 5000: $2000 < LD_{50} \leq 5000$ mg/kg
488   > 5000: $LD_{50} > 5000$ mg/kg
489   [2]The RC rat-only weight regression is log LD_{50} (mg/kg) = log IC_{50} (µg/mL) X 0.372 + 2.024.
490   [3]Reference rodent LD_{50} values from **Table 4-2**.
491   [4]Epinephrine bitartrate excluded because no rat LD_{50} was identified. Disulfoton and physostigmine excluded based on
492   mechanism of toxicity (see **Table 6-3**).
493   [5]Colchine excluded because no rat LD_{50} was identified. Endosulfan, parathion, potassium cyanide, and strychnine excluded based
494   on mechanism of toxicity (see **Table 6-3**).
495   [6]Dichlorvos, fenpropathrin, lindane, paraquat, phenobarbital, nicotine, and verapamil HCl excluded based on mechanism of
496   toxicity (see **Table 6-3**).
497   [7]Amitriptyline, atropine sulfate, caffeine, chloral hydrate, glutethimide, haloperidol, and procainamide HCl excluded based on
498   mechanism of toxicity (see **Table 6-3**).
499   [8]Carbon tetrachloride excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50}. Carbamazepine
500   excluded based on mechanism of toxicity (see **Table 6-3**).
501   [9]Methanol excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50}.
502   [10]Propylparaben excluded because no rat LD_{50} was identified.

503   • 4 (44%) of nine substances in the 2000 < LD_{50} ≤ 5000 mg/kg GHS toxicity

504   category were correctly predicted; however, since a total of 11 substances were

505   predicted for this category, the predictivity was 36% (4/11).

506   • 0 (0%) of 12 substances with LD_{50} > 5000 mg/kg were correctly predicted

507

508   The overall accuracy of the NHK NRU test method with the RC rat-only weight regression

509   was 30% [14 /47]) (**Table 6-5**). Toxicity was overpredicted for 47% (22) and underpredicted

510   for 23% (11) of the 47 substances, compared with *in vivo* toxicity categories (i.e., the GHS

511   categories for the reference LD_{50} values in **Table 4-2**). For this analysis, in terms of each

512   GHS toxicity classification category:

513   • 0 (0%) of four substances with LD_{50} < 5 mg/kg were correctly predicted

514   • 1 (14%) of seven substances in the 5 < LD_{50} ≤ 50 mg/kg GHS toxicity category

515   was correctly predicted

516   • 3 (60%) of five substances in the 50 < LD_{50} ≤ 300 mg/kg GHS toxicity category

517   were correctly predicted; however, since six other substances were also

518   predicted for this category, the predictivity was 33% (3/9)

519   • 8 (89%) of nine substances in the 300 < LD_{50} ≤ 2000 mg/kg GHS toxicity

520   category were predicted correctly; however, since 18 other substances were also

521   predicted for this category, the predictivity was 31% (8/26)

522   • 1 (11%) of nine substances in the 2000 < LD_{50} ≤ 5000 mg/kg GHS toxicity

523   category was correctly predicted

524   • 1 (8%) of 13 substances with LD_{50} > 5000 mg/kg was correctly predicted

525

526   *Discordant Substances for Prediction of Toxicity Category by the 3T3 and NHK NRU Test*

527   *Methods and the RC Rat-Only Weight Regression*

528   **Appendix L-2** shows the discordant substances for which the *in vitro* predicted GHS toxicity

529   category did not match that based on the reference rodent LD_{50} data using the RC rat-only

530   weight regression. The two *in vitro* NRU cytotoxicity test methods over- and under-

531   predicted the GHS toxicity category for a similar number of substances, compared with the

532   GHS toxicity categories for the reference LD_{50} values in **Table 4-2**. For the 3T3 NRU test

533   method, the GHS toxicity category of 19 (63%) of 30 discordant substances was

534   overpredicted and the GHS toxicity category of 11 (37%) substances was underpredicted.

535   For the NHK NRU test method, the GHS toxicity category of 22 (67%) of 33 discordant

536   substances was overpredicted and the toxicity of 11 (33%) discordant substances was

537   underpredicted.

538

539   6.3.3   Prediction of Toxicity Category by the 3T3 and NHK NRU Test Methods with the

540   RC Rat-Only Weight Regression Excluding Substances with Specific Mechanisms

541   of Toxicity

542   **Table 6-6** shows the concordance of the observed and predicted GHS acute oral toxicity

543   categories for each *in vitro* NRU test method using the geometric mean IC_{50} values (of the

544   three laboratories) and the RC rat-only weight regression after excluding substances with

545   specific mechanisms of toxicity (see **Table 6-3**). The formula for this regression was log

546   LD_{50} (mg/kg) = log IC_{50} (µg/mL) x 0.357 + 2.194. Accuracy is the agreement of the *in vitro*

547   predicted GHS toxicity categories with those based on the reference rat oral LD_{50} values

548   from **Table 4-2**.

549

550   *In Vitro – In Vivo Concordance for the 3T3 and NHK NRU Test Methods with the RC Rat-*

551   *Only Weight Regression Excluding Substances with Specific Mechanisms of Toxicity*

552   •   The overall accuracy of the 3T3 NRU test method with the RC rat-only weight

553   regression after excluding substances with specific mechanisms of toxicity was

554   46% (21/46 substances) (**Table 6-6**), compared to 35% (16/46 substances) when

555   the complete RC rat-only weight regression was used (**Section 6.3.2** and **Table**

556   **6-5**). **Table 6-6** shows that GHS toxicity category was overpredicted for 24%

557   (19) and underpredicted for 30% (11) of the 46 substances compared with the *in*

558   *vivo* GHS toxicity categories for the reference LD_{50} values in **Table 4-2**.

559

560   In terms of each GHS toxicity classification category:

561   •   0 (0%) of the four substances with LD_{50} < 5 mg/kg were correctly predicted

562   •   1 (14%) of seven substances in the $5 < LD_{50} \leq 50$ mg/kg GHS toxicity category

563   was correctly predicted

564

565   **Table 6-6   Prediction of GHS Toxicity Categories[1] by RC Rat-Only Weight Regression Excluding**

566   **Substances with Specific Mechanisms of Toxicity[2]**

Reference Rodent LD_{50}[3]	3T3 NRU-Predicted Toxicity Category						Total	Accuracy	Toxicity Overpredicted	Toxicity Underpredicted
	< 5	5 – 50	50 – 300	300-2000	2000-5000	> 5000				
< 5	0	0	2	2	0	0	4[4]	0%	100%	0%
5 – 50	0	1	4	2	0	0	7[5]	14%	86%	0%
50 – 300	0	0	4	1	0	0	5[6]	80%	20%	0%
300 – 2000	0	1	1	7	0	0	9[7]	78%	0%	22%
2000 – 5000	0	0	0	3	6	0	9[8]	67%	0%	33%
> 5000	0	0	0	5	4	3	12[9,10]	25%	0%	75%
Total	0	2	11	20	10	3	46	46%	24%	30%
Predictivity	0%	50%	36%	35%	60%	100%				
Category Underpredicted	0%	50%	9%	40%	40%	0%				
Category Overpredicted	0%	0%	55%	25%	0%	0%				
Reference Rodent LD_{50}[3]	NHK NRU-Predicted Toxicity Category						Total	Accuracy	Toxicity Overpredicted	Toxicity Underpredicted
	< 5	5 – 50	50 – 300	300 – 2000	2000 – 5000	> 5000				
< 5	0	0	2	2	0	0	4[4]	0%	100%	0%
5 – 50	0	1	4	2	0	0	7[5]	14%	86%	0%
50 – 300	0	1	3	1	0	0	5[6]	60%	20%	20%
300 – 2000	0	1	0	8	0	0	9[7]	89%	0%	11%
2000 – 5000	0	0	0	5	4	0	9[8]	44%	0%	56%
> 5000	0	0	0	4	7	2	13[10]	15%	0%	85%
Total	0	3	9	22	11	2	47	38%	23%	38%
Predictivity	0%	33%	33%	36%	36%	100%				
Category Underpredicted	0%	67%	0%	41%	64%	0%				
Category Overpredicted	0%	33%	67%	23%	0%	0%				

567   [1]Globally Harmonized System of Classification and Labelling of Chemicals with LD_{50} in mg/kg (UN 2005).
568

569 < 5:   LD$_{50}$ ≤ 5 mg/kg
570 5 – 50:   5 < LD$_{50}$ ≤ 50 mg/kg
571 50 – 300:   50 < LD$_{50}$ ≤ 300 mg/kg
572 300 – 2000:   300 < LD$_{50}$ ≤ 2000 mg/kg
573 2000 – 5000:   2000 < LD$_{50}$ ≤ 5000 mg/kg
574 > 5000:   LD$_{50}$ > 5000 mg/kg
575 [2]The RC rat-only weight regression excluding substances with specific mechanisms of toxicity is log LD$_{50}$ (mg/kg) = log IC$_{50}$
576 (μg/mL) X 0.357 + 2.194.
577 [3]Reference rodent LD$_{50}$ values from **Table 4-2**.
578 [4]Epinephrine bitartrate excluded because no rat LD$_{50}$ was identified. Disulfoton and physostigmine excluded based on
579 mechanism of toxicity (see **Table 6-3**).
580 [5]Colchine excluded because no rat LD$_{50}$ was identified. Endosulfan, parathion, potassium cyanide, and strychnine excluded based
581 on mechanism of toxicity (see **Table 6-3**).
582 [6]Dichlorvos, fenpropathrin, lindane, paraquat, phenobarbital, nicotine, and verapamil HCl excluded based on mechanism of
583 toxicity (see **Table 6-3**).
584 [7]Amitriptyline, atropine sulfate, caffeine, chloral hydrate, glutethimide, haloperidol, and procainamide HCl excluded based on
585 mechanism of toxicity (see **Table 6-3**).
586 [8]Carbon tetrachloride excluded because no laboratory attained sufficient toxicity for the calculation of an IC$_{50}$. Carbamazepine
587 excluded based on mechanism of toxicity (see **Table 6-3**).
588 [9]Methanol excluded because no laboratory attained sufficient toxicity for the calculation of an IC$_{50}$.
589 [10]Propylparaben excluded because no rat LD$_{50}$ was identified.
590

591   • 4 (80%) of five substances in the $50 < LD_{50} \leq 300$ mg/kg GHS toxicity category

592   were correctly predicted. Since seven other substances were also predicted for

593   this category, predictivity was 36% (4/11).

594   • 7 (78%) of nine substances in the $300 < LD_{50} \leq 2000$ mg/kg GHS toxicity

595   category were predicted correctly. Since a total of 20 substances were predicted

596   for this category, the predictivity was 35% (7/20).

597   • 6 (67%) of nine substances in the $2000 < LD_{50} \leq 5000$ mg/kg GHS toxicity

598   category were correctly predicted; the predictivity of this category was 60%

599   (6/10)

600   • 3 (25%) of 12 substances with $LD_{50} > 5000$ mg/kg were correctly predicted.

601   Since no other substances were predicted for this category, the predictivity was

602   100% (3/3).

603

604   **Table 6-6** shows that the accuracy of the NHK NRU test method with the RC rat-only weight

605   regression excluding substances with specific mechanisms of toxicity was 38% (18/47),

606   compared to the 30% (14/47) accuracy when the complete RC rat-only weight regression was

607   used (see **Table 6-5**). Toxicity was overpredicted for 23% (11) and underpredicted for 38%

608   (19) of the 47 substances compared with the *in vivo* GHS categories for the reference LD_{50}

609   values in **Table 4-2**. In terms of each GHS toxicity classification category:

610   • 0 (0%) of the four substances with $LD_{50} < 5$ mg/kg were correctly predicted

611   • 1 (14%) of seven substances in the $5 < LD_{50} \leq 50$ mg/kg GHS toxicity category

612   was correctly predicted

613   • 3 (60%) of five substances in the $50 < LD_{50} \leq 300$ mg/kg GHS toxicity category

614   were correctly predicted. Since six other substances were also predicted for this

615   category, predictivity was 33% (3/9).

616   • 8 (89%) of nine substances in the $300 < LD_{50} \leq 2000$ mg/kg GHS toxicity

617   category were predicted correctly. Since 14 other substances that did not match

618   this category were also predicted, predictivity was 36% (8/22).

619   • 4 (44%) of nine substances in the $2000 < LD_{50} \leq 5000$ mg/kg GHS toxicity

620   category were correctly predicted; the predictivity of this category was 36%

621   (4/11)

622   •   2 (15%) of 13 substances with LD_{50} > 5000 mg/kg were correctly predicted.

623   Since no other substances were predicted for this category, the predictivity was

624   100% (2/2).

625

626 *Discordant Substances for the Prediction of Toxicity Category by the 3T3 and NHK NRU*

627 *Test Methods and the RC Rat-Only Weight Regression Excluding Substances with Specific*

628 *Mechanisms of Toxicity*

629 **Appendix L-2** shows the discordant substances for which the *in vitro* NRU predicted toxicity

630 category did not match that based on the reference rodent LD_{50} data. The NHK NRU test

631 method had four more discordant substances than the corresponding assay using 3T3 cells

632 when the IC_{50} results were applied to the RC rat-only weight regression excluding substances

633 with specific mechanisms of toxicity. For the 3T3 NRU test method, the GHS toxicity

634 category of 19 (63%) of 30 discordant substances was overpredicted while the toxicity of 11

635 (37%) of 30 discordant substances was underpredicted compared with the *in vivo* GHS

636 toxicity categories for the reference LD_{50} values in **Table 4-2**. For the NHK NRU test

637 method, the toxicity of 22 (65%) of 34 discordant substances was overpredicted while the

638 toxicity of 12 (35%) of 34 discordant substances was underpredicted.

639

640 6.3.4   <u>Summary of the Regressions Evaluated</u>

641 **Table 6-7** summarizes the regressions evaluated in **Section 6.3** for accuracy in predicting the

642 GHS acute oral toxicity categories (UN 2005) and the proportion of discordant substances for

643 *in vitro* predictions of GHS toxicity categories. Accuracy for the 3T3 NRU test method was

644 slightly lower than that for the NHK NRU test method for the RC millimole regression (i.e.,

645 26% vs. 28%). Accuracy for the 3T3 NRU test method was higher than that for the NHK

646 NRU test method for the RC rat-only weight regression (i.e., 35% vs. 30%) and the RC rat-

647 only weight regression excluding substances with specific mechanisms of toxicity (i.e., 46%

648 vs. 38%). The proportion of discordant substances for the 3T3 NRU test method was higher

649 for the RC millimole regression (74%) than it was for the RC rat-only weight (65%)

650 regression and the RC rat-only weight regression excluding substances with specific

651 mechanisms of toxicity (65%). The proportion of discordant substances for the NHK NRU

652 test method was similar for each regression (i.e., 70-72%). **Table 6-7** shows that the

653   difference between the proportions of discordant substances for the 3T3 and NHK NRU test

654   methods widened with each subsequent regression (74% vs. 72% for the RC millimole

655   regression, 65% vs. 70% for the RC rat-only weight regression, and 65% vs. 72% for the RC

656   rat-only weight regression excluding substances with specific mechanisms of toxicity).

657

658   **Table 6-7   Comparison of Regressions and *In Vitro* NRU Test Methods for**

659   **Performance in Predicting GHS[a] Toxicity Categories**

Regression	N[b]	Adjusted R^2	Accuracy	Discordant Substances[c]
RC –millimole units	347	0.450[d]	3T3 – 26% NHK – 28%	3T3 – 34/46 (74%) NHK – 34/47 (72%)
RC rat only –weight units[e]	282	0.322	3T3 – 35% NHK – 30%	3T3 – 30/46 (65%) NHK – 33/47 (70%)
RC rat only excluding substances with specific mechanisms of action – weight units[e]	232	0.353	3T3 – 46% NHK – 38%	3T3 – 30/46 (65%) NHK – 34/47 (72%)

660   [a]Globally Harmonized System of Classification and Labelling of Chemicals with LD_{50} in mg/kg (UN 2005).
661   [b]Number of substances used in regression.
662   [c]Proportion of substances evaluated.
663   [d]Calculated from RC data (i.e., regression not reported by Halle [1998]).
664   [e]From **Table 6-1**.
665

666

667   The highest accuracy for both *in vitro* NRU cytotoxicity test methods was attained when

668   using the RC rat-only weight regression excluding substances with specific mechanisms of

669   toxicity. The accuracy for the 3T3 NRU test method was 46%, which was greater than the

670   accuracy of the 3T3 NRU with the RC millimole regression (26%) and with the RC rat-only

671   weight regression (35%). The accuracy for the NHK NRU test method was 38% for the RC

672   rat-only weight regression excluding substances with specific mechanisms of toxicity, 28%

673   with the RC millimole regression, and 30% with the RC rat-only weight regression.

674

675   **6.4   Strengths and Limitations of the *In Vitro* NRU Cytotoxicity Test Methods for**

676   ***In Vivo* Toxicity Prediction**

677

678   For each regression evaluated, the NRU basal cytotoxicity test methods tended to

679   underpredict the toxicity of the most toxic substances and to overpredict the toxicity of the

680   least toxic substances. The 3T3 and NHK NRU test methods were better at predicting the

681   toxicity of substances with $50 < LD_{50} \leq 300$ mg/kg and $300 < LD_{50} \leq 2000$ mg/kg than

682   predicting the toxicity of substances with higher or lower LD_{50} values. The accuracy for the

683   RC millimole regression and the RC rat-only weight regression for these toxicity categories

684   was 67-100% for the 3T3 NRU and 33-83% for the NHK NRU data. Substances with higher

685   or lower LD_{50} values were infrequently predicted correctly. The accuracy for substances

686   with $LD_{50} \leq 50$ mg/kg (GHS toxicity categories for $LD_{50} \leq 5$ mg/kg and $5 < LD_{50} \leq 50$

687   mg/kg) was 0-17% for the 3T3 NRU and 0-50% for the NHK NRU with the same

688   regressions. Accuracy for substances in the $2000 < LD_{50} \leq 5000$ and $LD_{50} > 5000$ mg/kg

689   toxicity categories was 0-44% for the 3T3 NRU and 0-11% for the NHK NRU.

690

691   The RC rat-only weight regression calculated after removal of substances with specific

692   mechanisms of toxicity improved the accuracy of GHS toxicity category predictions for

693   substances with $LD_{50} > 2000$ mg/kg compared with the accuracy for the other regressions.

694   The accuracy for substances in these categories was 25-67% for the 3T3 NRU and 15-44%

695   for the NHK NRU. The RC rat-only weight regression excluding substances with specific

696   mechanisms of toxicity did not increase the accuracy for substances with $LD_{50} < 2000$

697   mg/kg. However, the accuracy for substances in the $50 < LD_{50} \leq 300$ mg/kg and $300 < LD_{50}$

698   ≤ 2000 mg/kg categories using the RC millimole regression and the RC rat-only weight

699   regression was already quite high. The accuracy for predicting these categories using the RC

700   rat-only weight regression excluding substances with specific mechanisms of toxicity was

701   78-80% for the 3T3 NRU and 60-89% for the NHK NRU. The accuracy for predicting the

702   toxicity categories for $LD_{50} \leq 5$ mg/kg and $5 < LD_{50} \leq 50$ mg/kg was 0-14% for both the 3T3

703   NRU test methods when using the RC rat-only weight regression excluding substances with

704   specific mechanisms of toxicity.

705

706   The analysis of the 30 (3T3 NRU) to 31 (NHK NRU) discordant substances for the RC

707   millimole regression to determine the physical, chemical, and biological characteristics

708   associated with the discordant substances is presented in **Appendix L-1**. The analysis

709   showed that 3 of 3 (100%) organophosphates were discordant in both test methods (10% of

710   the 30 [3T3 NRU)] to 31 [NHK NRU] discordant substances). Other characteristics that

711   seemed promising for characterizing RC millimole regression outliers were boiling point,

712   molecular weight, and log K_{ow}. For boiling points > 200°C, 9/13 substances (69%) were

713   outliers for both the 3T3 results NHK NRU results (29 and 26% of the outliers, respectively).

714   The toxicity of seven of the nine (78%) outliers with boiling points > 200°C was

715   underpredicted by the RC millimole regression and the toxicity of the other two (22%)

716   substances was overpredicted. For molecular weight > 400 g/mole, 5/7 (71%) substances

717   were outliers using the 3T3 data and 3/7 (43%) were outliers using the NHK data (17 and

718   10% of the outliers, respectively). The toxicity of all the outliers with molecular weight >

719   400 g/mole was underpredicted by the RC millimole regression (5/5 [100%] for the 3T3

720   NRU and 3/3 [100%] for the NHK NRU). For log K_{ow} > 3, 9/12 (75%) substances were

721   outliers using the 3T3 data (30% of the outliers) and 8/12 (67%) substances were outliers

722   using the NHK data (26% of the outliers). The toxicity of 7/9 (78%) outliers (with

723   log K_{ow} > 3) for the 3T3 NRU assay was underpredicted by the RC millimole regression and

724   the toxicity of 2/9 (22%) outliers was overpredicted. The toxicity of 6/8 (75%) outliers (with

725   log K_{ow} > 3) for the NHK NRU assay was underpredicted by the RC millimole regression

726   and the toxicity of 2/8 (25%) outliers was overpredicted. Of the 21 substances with specific

727   mechanisms of toxicity that were not expected to be active in the 3T3 and NHK cell cultures,

728   13 (62%) were outliers. These substances represented 13/30 (43%) of the discordant

729   substances for the 3T3 NRU and 13/31 (42%) for the NHK NRU.

730

731   The lack of fit of individual substances to the regressions was not consistently related to their

732   insolubility in the 3T3 or NHK medium. Of the 25 substances that exhibited precipitates in

733   the 3T3 NRU assay, 11 (44%) substances were discordant (see **Table 5-8** for substances that

734   had precipitates and **Appendix L-1** for the analysis of discordant substances). The toxicity

735   of nine of the 11 (82%) substances was underpredicted by the RC millimole regression and

736   the toxicity of two of the 11 (18%) substances was overpredicted by the RC millimole

737   regression. Of the 24 substances that exhibited precipitates in the NHK NRU assay, 11

738   (46%) substances were outliers. The toxicity of nine of the 11 (82%) substances was

739   underpredicted by the RC millimole regression and the toxicity of two of the 11 (18%)

740   substances was overpredicted by the RC millimole regression.

741

742   Additionally, the lack of fit of individual substances to the RC millimole regression was not

743   consistently related to the fact that the test method systems had little to no metabolic

744   capability. Such a system would be expected to underestimate the toxicity of substances with

745   active metabolites. However, the toxicity of substances known to produce active metabolites

746   *in vivo* (listed in **Table 3-7**) was not necessarily underpredicted by the NRU assays. Of the

747   19 substances known to produce active metabolites *in vivo*, ten were discordant in the 3T3

748   NRU test method. Of these ten discordant substances, the toxicity of six (60%) was

749   underpredicted while the toxicity of four (40%) was overpredicted by the 3T3 NRU test

750   method. These ten discordant substances accounted for 33% of the 30 discordant substances

751   identified for the 3T3 NRU test method. Nine of the 19 substances known to produce active

752   metabolites *in vivo* were discordant for the NHK NRU test method. Of these nine discordant

753   substances, the NHK NRU assay underpredicted the toxicity of five (56%) substances and

754   overpredicted the toxicity of four (44%) substances. These nine discordant substances

755   accounted for 29% of the 31 discordant substances identified for the NHK NRU test method.

756

757   Similarly, Halle (1998) noted that the RC substances that required metabolic activation to

758   produce *in vivo* toxicity were not necessarily discordant substances (with respect to fit to the

759   RC millimole regression). Halle (1998) found that eight (50%) of the 16 substances that

760   required metabolic activation to product toxicity were discordant substances while eight

761   (50%) were not discordant (see **Table L3-3** in **Appendix L3**).

762

763   Some substances with low toxicity and low solubility could not be tested in the *in vitro* NRU

764   cytotoxicity assays because the amount of dissolved substance was inadequate to obtain an

765   IC$_{50}$ value. In the 3T3 NRU test method, none of the laboratories obtained adequate toxicity

766   in any experiment with carbon tetrachloride and methanol. At least one laboratory failed to

767   achieve adequate toxicity with gibberellic acid and xylene. In the corresponding NHK assay,

768   no laboratory achieved adequate toxicity in any experiment with carbon tetrachloride, and at

769   least one laboratory could not achieve adequate toxicity with methanol, 1,1,1-trichloroethane,

770   and xylene.

771

772   Although the accuracy of the 3T3 and NHK NRU test methods for predicting *in vivo* toxicity

773   category was rather low when used with the RC millimole regression and the RC rat-only

774   weight regression, it was improved by removing substances with specific mechanisms of

775   toxicity that were not expected to be active in the 3T3 and NHK cell cultures. The evaluation

776   of these *in vitro* NRU cytotoxicity test methods for predicting starting doses for acute

777   systemic toxicity testing, thereby reducing and refining animal use, is provided in **Section 10**.

778

779   **6.5   Salient Issues of Data Interpretation**

780

781   One of the most important considerations for the 3T3 and NHK NRU test methods is getting

782   good dose-response results. In addition to technical difficulties with these methods, such as

783   occasional poor cell growth and the formation of NRU crystals, this validation study yielded

784   observations of unusual dose-responses for certain substances.

785

786   The experimenter must be aware of dose-response anomalies and their causes in order to

787   determine whether the dose-response can be better defined. For example, for substances

788   such as aminopterin, which generally produced a biphasic dose-response using the log-dose

789   spacing of the range-finder test, the experimenter must focus on the lowest concentration at

790   which the substance produced 50% toxicity in order to perform the definitive testing with

791   more closely spaced concentrations. In the definitive tests of such substances, the toxic

792   response may plateau before producing 100% toxicity (i.e., 0% viability). The method used

793   for the calculation of the IC_{50} must reflect an IC_{50} that is 50% inhibition of the control values

794   rather than the midpoint of the highest and lowest response (as provided by the standard Hill

795   function analysis).

796

797   Some substances, because of their low toxicity and/or low solubility, do not provide

798   sufficient toxicity for the calculation of an IC_{50} value. Carbon tetrachloride, methanol,

799   xylene, gibberellic acid, lithium carbonate and 1,1,1-trichloroethane failed to yield acceptable

800   IC_{50} results in at least one laboratory due to insufficient toxicity/insolubility. All of these

801   substances, with the exception of methanol, were reported to produce precipitate in the cell

802   culture medium.

803

804

804   **6.6   Comparison to Established Performance Standards**

805

806   The *Guidance Document* method of evaluating basal cytotoxicity assays for use in predicting

807   starting doses for acute oral toxicity assays provides the existing performance standard

808   (ICCVAM 2001b) for the 3T3 and NHK NRU cytotoxicity test methods. The *Guidance*

809   *Document* recommends testing10 to 20 reference substances from the RC in a candidate *in*

810   *vitro* basal cytotoxicity assay to be used for predicting starting doses (ICCVAM 2001b). The

811   substances should cover a wide range of toxicity and fit the RC prediction model (i.e., the

812   linear regression line) as closely as possible. The IC_{50} results for the selected reference

813   substances are used to calculate a new regression line with the LD_{50} values used by the RC.

814   If the resulting regression is parallel to the RC millimole regression and within the ± log 5

815   (i.e., ± 0.699) prediction interval for the RC, the *Guidance Document* recommends using the

816   cytotoxicity assay to predict starting doses for unknown substances to be tested in acute oral

817   systemic toxicity assays.

818

819   One goal of the coded substance testing in Phases Ib and II of this study was to establish

820   whether the results from the 3T3 and NHK NRU cytotoxicity test methods were consistent

821   with the RC millimole regression. As discussed in **Section 3.4.1**, two of the major criteria

822   for selecting the 12 coded substances tested in these phases from the 72 substances to be

823   tested were (a) two substances must be included from each of the unclassified and classified

824   GHS acute oral toxicity categories and (b) the substances must fit as closely to the RC

825   millimole regression as possible. Unfortunately, the SMT could not identify 12 substances

826   that fit both criteria since there was only one substance, aminopterin, in the $LD_{50} < 5$ mg/kg

827   category that fit the RC millimole regression. The other substance chosen for testing for that

828   toxicity category was sodium selenate. Since sodium selenate was not included in the RC,

829   the SMT did not know how closely it would fit the RC millimole regression and it was not

830   included in the regression analyses for Phases Ib and II.

831

832   The geometric mean log IC_{50} values from the 3T3 and NHK test methods from each

833   laboratory were used with the oral log LD_{50} values from the RC (see **Appendices J-1 and J-**

834   **3**) for the calculation of least squares linear regression analyses (see **Section 5.3**) for the

835 substances tested in Phases Ib and II. The slopes for all regressions were significantly

836 different from zero with p < 0.0001. The adjusted R^2 values for the regressions from each

837 laboratory, shown in **Table 6-8**, indicate that the 3T3 NRU test method produced better

838 fitting regressions than the corresponding NHK assay (adjusted R^2 = 0.934 – 0.947 vs. 0.530

839 – 0.579). The relatively low adjusted R^2 values for the NHK assay were attributed to the

840 much lower toxicity of aminopterin in that assay (see **Figures 6-6** to **6-8** and **Table 5-4**).

841 The regressions were consistent with the RC millimole regression. **Table 6-8** shows that

842 p > 0.01, the level of statistical significance, for all joint comparisons of slopes and intercepts

843 with the RC millimole regression. The RC millimole regression slope and intercept were

844 assumed to be constants for the comparison. A graphical comparison of the regressions with

845 the RC millimole regression as suggested by the *Guidance Document* (ICCVAM 2001b)

846 examples demonstrated that the regressions were generally within the RC millimole

847 regression acceptance limits (see **Figures 6-6** to **6-8**). According to the *Guidance Document*

848 (ICCVAM 2001b), basal cytotoxicity assays providing such consistency with the RC

849 millimole regression are acceptable for predicting starting doses for *in vivo* acute oral

850 systemic toxicity assays.

851

852 **Table 6-8   Linear Regressions for Substances Tested in Phases Ib and II**

	3T3 Millimole Regression			P-Values for Test Against RC Millimole Regression		
Laboratory	Intercept	Slope	Adjusted R^2	Intercept	Slope	Joint[1]
ECBC	0.793	0.584	0.934	0.202	0.014	0.040
FAL	0.709	0.598	0.947	0.497	0.008	0.024
IIVS	0.710	0.584	0.943	0.508	0.014	0.041
	NHK Millimole Regression			P-Values for Test Against RC Millimole Regression		
Laboratory	Intercept	Slope	Adjusted R^2	Intercept	Slope	Joint[1]
ECBC	0.401	0.530	0.530	0.484	0.547	0.620
FAL	0.429	0.548	0.579	0.519	0.450	0.569
IIVS	0.373	0.549	0.544	0.426	0.475	0.538

853 [1]Simultaneous comparison of slope and intercept. The RC slope and intercept were assumed to be constants.
854 ECBC – US Army Edgewood Chemical Biological Center; FAL – FRAME Alternatives Laboratory; IIVS –
855 Institute for *In Vitro* Sciences
856

857

858

858   **Figure 6-6**   *In Vitro – In Vivo* **Regressions[1] for Phases Ib and II for ECBC**
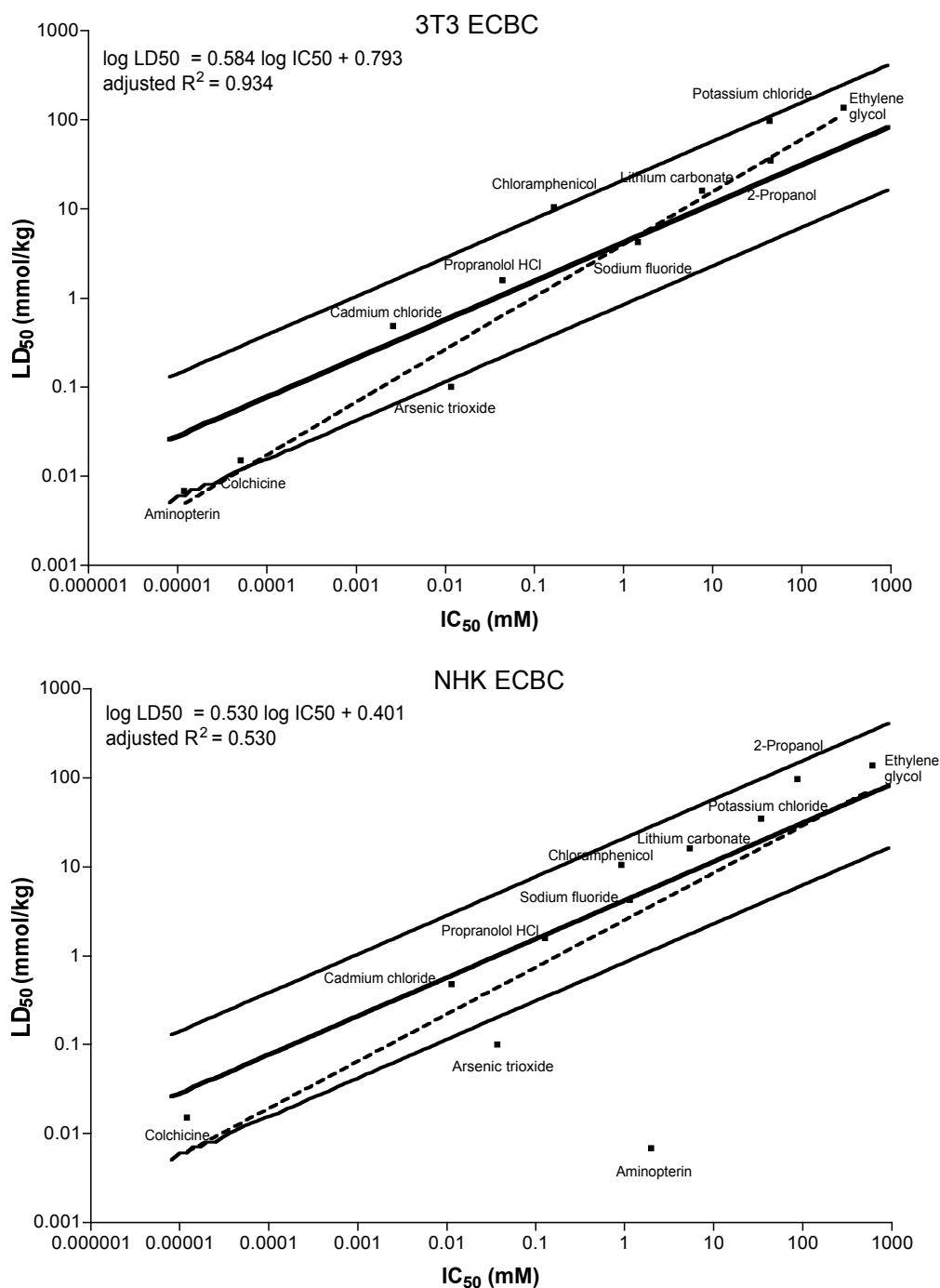
859

860   —— Solid Lines: RC millimole regression and acceptance limits   - - - - - Dashed Line: Study Regression

861   [1]Regressions of substances tested in Phases Ib and II do not include sodium selenate since it was not included in

862   the RC.

863   ECBC: U.S. Army Edgewood Chemical Biological Command

864

865   **Figure 6-7   *In Vitro – In Vivo* Regressions[1] for Phases Ib and II for FAL**
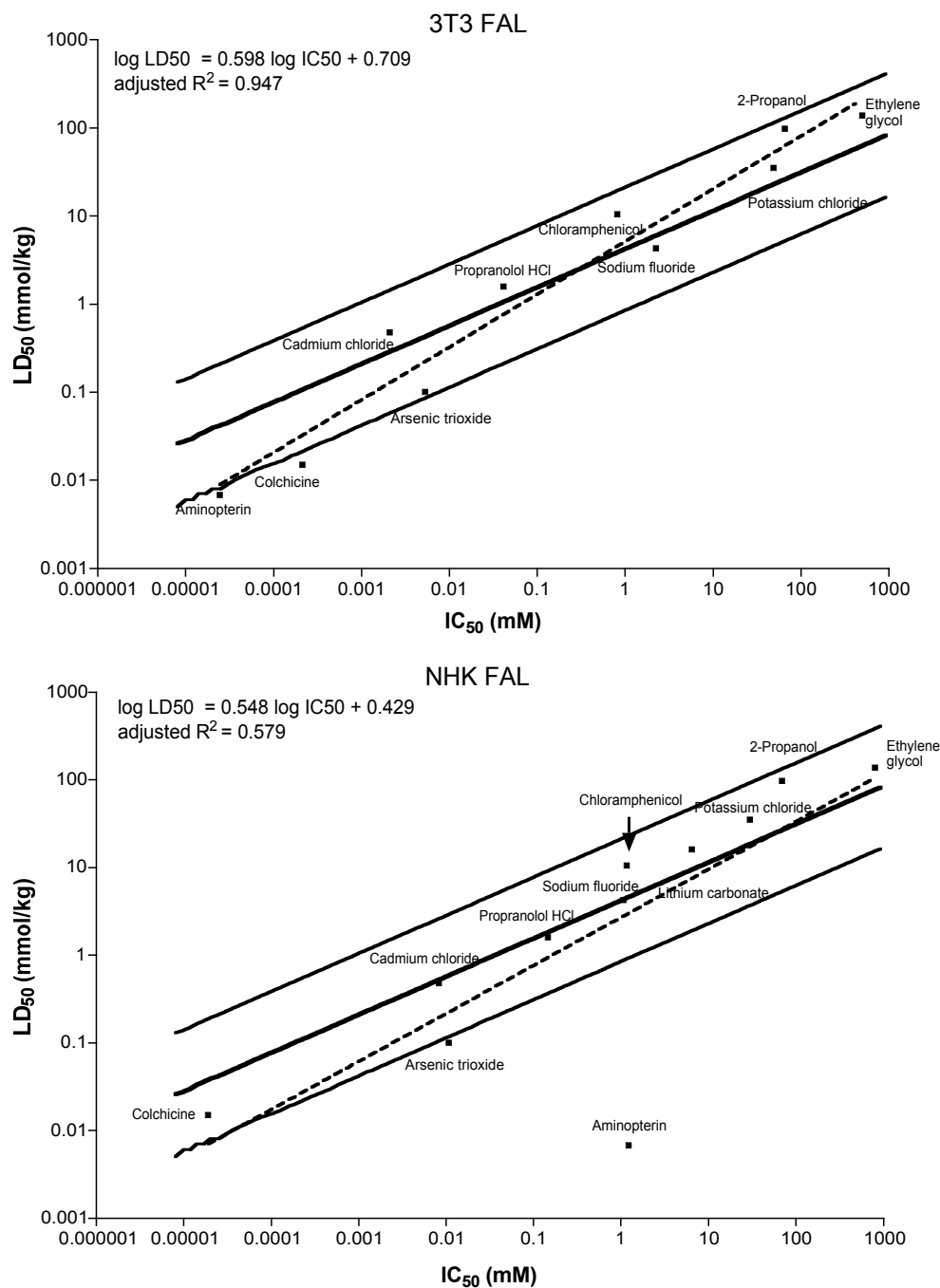
866
867   —— Solid Lines: RC millimole regression and acceptance limits   - - - - - Dashed Line: Study Regression
868   [1]Regressions of substances tested in Phases Ib and II do not include sodium selenate since it was not included in
869   the RC.
870   FAL: Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory
871

872

872   **Figure 6-8   *In Vitro – In Vivo* Regressions[1] for Phases Ib and II for IIVS**
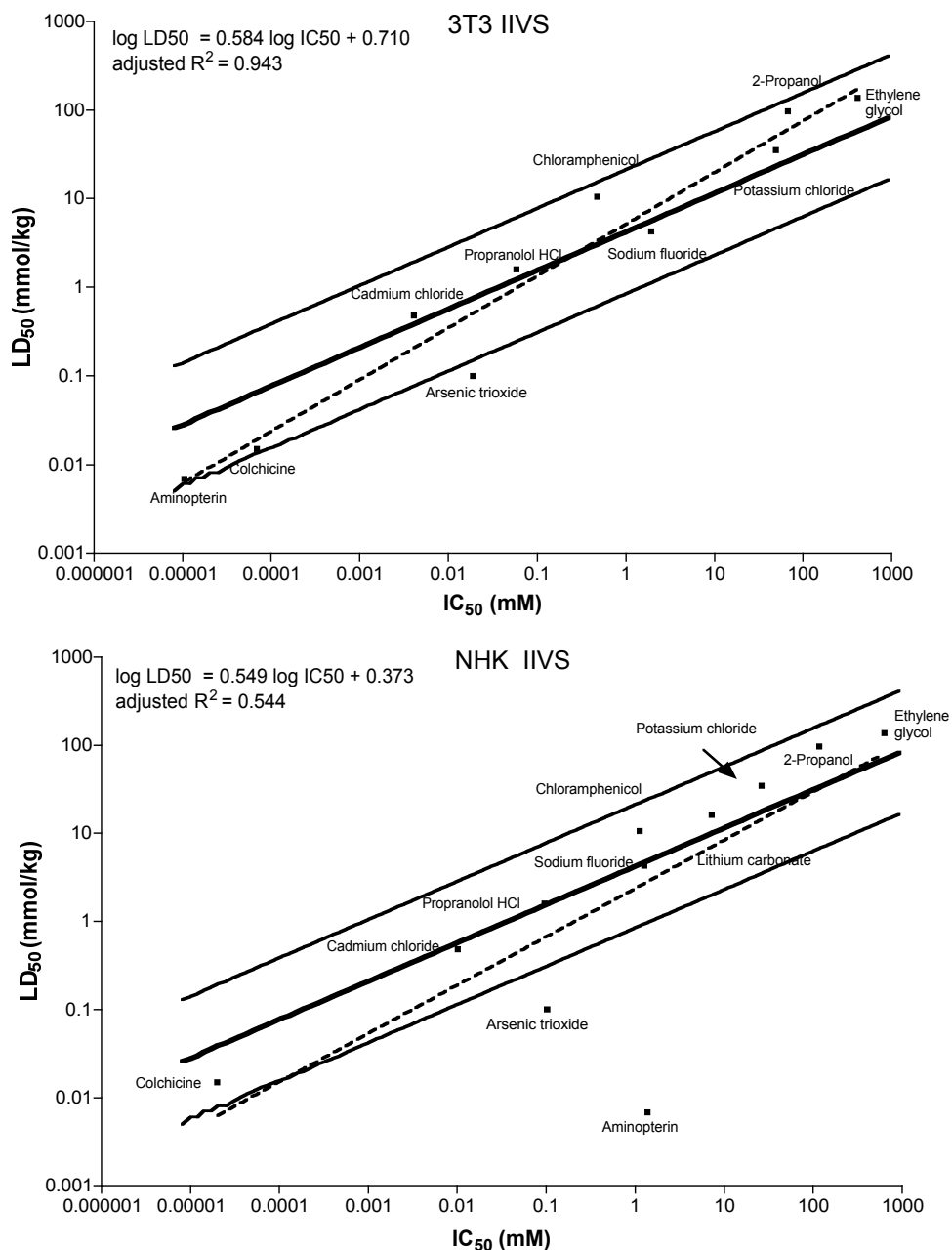
873

874   —— Solid Lines: RC millimole regression and acceptance limits   - - - - - Dashed Line: Study Regression
875   [1]Regressions of substances tested in Phases Ib and II do not include sodium selenate since it was not included in
876   the RC.
877   IIVS: Institute for In Vitro Sciences

878

879

880

881   **6.7   Summary**

882

883   The millimole regressions developed using the NICEATM/ECVAM IC_{50} and LD_{50} data were

884   not significantly different from a regression for the 58 RC substances calculated using the RC

885   data (F test; p = 0.929 for the 3T3 NRU regression and p = 0.144 for the NHK NRU

886   regression). To improve the RC millimole regression with respect to the prediction of LD_{50}

887   values by *in vitro* NRU IC_{50} values, regressions were developed using the RC data in weight

888   units to exclude (1) mouse data (i.e., the RC rat-only weight regression) and (2) substances

889   with mechanisms of toxicity that were not expected to be active in the 3T3 and NHK cell

890   cultures (i.e., the RC rat-only regression excluding substances with specific mechanisms of

891   toxicity regression).

892

893   Accuracy in predicting GHS acute toxicity category using these *in vitro* NRU cytotoxicity

894   test methods was 26% (12/46) for the 3T3 NRU and 28% (13/47) for the NHK NRU with the

895   RC millimole regression. Accuracy with the RC rat-only weight regression improved to 35%

896   (16/46) for the 3T3 NRU and 30% (14/47) for the NHK NRU. Accuracy was higher for

897   substances with $50 < LD_{50} \le 2000$ mg/kg compared to substances with higher or lower

898   toxicity. For these two regressions, the accuracy of predicting the $50 < LD_{50} \le 300$ mg/kg

899   and $300 < LD_{50} \le 2000$ mg/kg categories for the 3T3 and NHK NRU was 67-100% and 50-

900   100%, respectively. The accuracy of predicting the $LD_{50} \le 5$ mg/kg and $5 < LD_{50} \le 50$

901   mg/kg categories was 0-17% for the 3T3 NRU and 0-50% for the NHK NRU. The accuracy

902   for substances with $2000 < LD_{50} \le 5000$ mg/kg and $LD_{50} > 5000$ mg/kg was 0-67% and

903   0-44% for the 3T3 and NHK NRU data, respectively.

904

905   Examination of outliers for the RC millimole regression by chemical class showed that 3 of 3

906   (100%) organophosphates were outliers in both test methods. Other characteristics that

907   seemed promising for characterizing RC outliers were boiling point, molecular weight, and

908   log K_{ow}. For boiling points $> 200°C$, 9/13 (69%) substances were outliers for both the 3T3

909   and NHK NRU results. For molecular weight > 400 g/mole, 5/7 (71%) substances were

910

910   outliers using the 3T3 data and 3/7 (43%) were outliers using the NHK data. For log K_{ow} >

911   3, 9/12 (75%) substances were outliers using the 3T3 data and 8/12 (67%) substances were

912   outliers using the NHK data.

913

914   The lack of fit of individual substances to the RC millimole regression was not consistently

915   related to substance insolubility in the 3T3 or NHK medium or to the fact that the test

916   method systems had little to no metabolic capability. Of the substances that exhibited

917   precipitates, 11/25 (44%) substances were discordant with the 3T3 NRU assay and 11/24

918   (46%) were discordant with the NHK NRU assay. Also, although the 3T3 and NHK cells

919   have little to no metabolic capability, the toxicity of substances known to produce active

920   metabolites *in vivo* was not necessarily underpredicted by these assays. Of the 19 substances

921   known to produce active metabolites *in vivo*, ten (53%) were discordant in the 3T3 NRU test

922   method. Of these ten discordant substances, the toxicity of six (60%) was underpredicted

923   while the toxicity of four (40%) was overpredicted by the 3T3 NRU test method. These ten

924   discordant substances accounted for 33% of the 30 discordant substances identified for the

925   3T3 NRU test method. Similarly, nine (47%) of the 19 substances known to produce active

926   metabolites *in vivo* were discordant for the NHK NRU test method. Of these nine discordant

927   substances, the NHK NRU assay underpredicted the toxicity of five (56%) substances and

928   overpredicted the toxicity of four (44%) substances. These nine discordant substances

929   accounted for 29% of the 31 discordant substances identified for the NHK NRU test method.

930

931   The examination of outliers based on mechanism of toxicity lead to the development the RC

932   rat-only weight regression excluding substances with specific mechanisms of toxicity. Of the

933   21 substances with specific mechanisms of toxicity that were not expected to be active in the

934   3T3 and NHK cell cultures, 13 (62%) were outliers. These substances represented 13/30

935   (43%) of the discordant substances for the 3T3 NRU test method and 13/31 (42%) for the

936   NHK NRU test method. The RC rat-only weight regression excluding substances with

937   specific mechanisms of toxicity improved the accuracy from 26% (12/46) for the RC

938   millimole regression to 46% (21/46) for the 3T3 NRU test method and from 28% (13/47) to

939   38% (18/47) for the NHK NRU test method.

940

941   The RC rat-only weight regression calculated after removal of substances with specific

942   mechanisms of toxicity improved the accuracy (compared with the RC millimole regression)

943   for predicting most toxicity categories. It did not improve the accuracy of category

944   prediction for substances with LD_{50} < 50 mg/kg or for substances with 300 < LD_{50} ≤ 2000

945   mg/kg. The following changes in accuracy for the various toxicity categories, compared with

946   the RC millimole regression, occurred:

947   • LD_{50} ≤ 5 mg/kg– 0% to 0% for both 3T3 and NHK NRU

948   • 5 < LD_{50} ≤ 50 mg/kg – 17% to 14% for the 3T3 NRU and 50% to 14% for the

949   NHK NRU

950   • 300 < LD_{50} ≤ 2000 mg/kg – 100% to 78% for the 3T3 NRU and 100% to 89%

951   for the NHK NRU

952   • 2000 < LD_{50} ≤ 5000 mg/kg – 0% to 67% for the 3T3 NRU and 9% to 44% for

953   the NHK NRU

954   • LD_{50} > 5000 mg/kg – 10% to 25% for the 3T3 NRU and 0% to 15% for the

955   NHK NRU

956

956
957
958
959
960
961
962
963
964
965
966
967
968
969
970   ***[This Page Intentionally Left Blank]***
971
972

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39   ***[This Page Intentionally Left Blank]***

40

41

41   **7.0   RELIABILITY OF THE 3T3 AND NHK NRU TEST METHODS**

42

43   This section discusses the reliability of the 3T3 and NHK NRU test methods. Reliability is

44   the degree to which a test method can be performed reproducibly within and among

45   laboratories over time (ICCVAM 2003). It is assessed by calculating intra- and inter-

46   laboratory reproducibility and repeatability. Reproducibility is the consistency of individual

47   test results obtained in a single laboratory (intralaboratory reproducibility) or in different

48   laboratories (interlaboratory reproducibility) using the same protocol and test samples.

49   Repeatability, usually applied to results within a laboratory, is the closeness of agreement

50   between test results obtained within a single laboratory when the procedure is performed on

51   the same substance under identical conditions within a given time. The NICEATM/ECVAM

52   study was not designed to assess intralaboratory repeatability.

53

54   For the NICEATM/ECVAM validation study, reliability was assessed by determining both

55   intra- and inter-laboratory reproducibility. Intralaboratory reproducibility is the agreement of

56   results produced when qualified people within the same laboratory perform the test method

57   using the same test protocol at different times (ICCVAM 2003). Interlaboratory

58   reproducibility is the agreement of results from different qualified laboratories using the

59   same protocol and reference substances. Interlaboratory reproducibility indicates the extent

60   to which a test method can be successfully transferred among laboratories.

61

62   Intra- and inter-laboratory reproducibility of the 3T3 and NHK NRU test methods were

63   determined using ANOVA and CV analysis as discussed in **Section 5.3.3 (**see **Sections 7.2.1**

64   and **7.2.2**). Interlaboratory reproducibility of the 3T3 and NHK NRU test methods was also

65   assessed by comparing the laboratory-specific IC_{50}-LD_{50} regressions (from **Table 6-1**) to one

66   another for each test method (see **Section 7.2.3**) and by evaluating laboratory concordance

67   for the GHS acute oral toxicity category predictions reported in **Sections 6.3.1** through **6.3.3**

68   (see **Section 7.2.4**). Laboratory concordance for the solvent selection process using the

69   solubility protocol (described in **Section 2.9**) is provided in **Section 7.4**.

70

71

72 **7.1   Substances Used to Determine the Reliability of the 3T3 and NHK NRU Test**

73 **Methods**

74

75 The SMT intended to use the IC_{50} results of all 72 reference substances identified for testing

76 in **Table 3-2** to determine the reliability of the 3T3 and NHK NRU test methods.

77 Unfortunately, IC_{50} results for all substances could not be obtained in all the laboratories.

78 **Table 7-1** shows the substances that failed to yield sufficient cytotoxicity for the calculation

79 of an IC_{50} and the number of substances left to determine intralaboratory reproducibility. The

80 laboratories failed to obtain IC_{50} results for three to five substances in the 3T3 NRU test

81 method and two to three substances with the NHK NRU test method.

82

83 For the 3T3 NRU test method, no laboratory achieved sufficient cytotoxicity to obtain IC_{50}

84 values for carbon tetrachloride or methanol and only one laboratory obtained IC_{50} results for

85 lithium carbonate and xylene. Thus, interlaboratory reproducibility for the 3T3 NRU test

86 method was assessed using the remaining 68 reference substances. For the NHK NRU test

87 method, no laboratory obtained IC_{50} values for carbon tetrachloride and only one laboratory

88 achieved IC_{50} results for xylene and 1,1,1-trichloroethane. Interlaboratory reproducibility for

89 the NHK NRU test method was assessed using the IC_{50} results for the remaining 69 reference

90 substances.

91

92 Despite the fact that IC_{50} values were not obtained by all the laboratories for all reference

93 substances, **Table 7-2** shows that the complete range of LD_{50} responses, as defined by the

94 GHS classification for acute oral toxicity in **Table 3-1**, was covered by the remaining

95 substances. The IC_{50} values also covered a wide range of responses (see **Table 7-3**). IC_{50}

96 values for the 3T3 NRU test method ranged from 0.005 to 38,878 μg/mL. IC_{50} values for the

97 NHK NRU test method covered a larger range, from 0.00005 to 49,800 μg/mL.

98

99

100

100　**Table 7-1　　Reference Substances That Failed to Yield IC$_{50}$ Values[1] And Number of**

101　　　　　　　　**Reference Substances Available for Intralaboratory Reproducibility**

102　　　　　　　　**Analyses**

Laboratory	3T3 NRU Test Method		NHK NRU Test Method	
	Reference Substances Lacking IC$_{50}$ Results	**N[2]**	**Reference Substances Lacking IC$_{50}$ Results**	**N[2]**
ECBC	Carbon tetrachloride Methanol Xylene	69	Carbon tetrachloride Methanol Xylene	69
FAL	Carbon tetrachloride Gibberellic acid Lithium carbonate Methanol Xylene	67	1,1,1-Trichloroethane Carbon tetrachloride Xylene	69
IIVS	Carbon tetrachloride Lithium carbonate Methanol	69	1,1,1-Trichloroethane Carbon tetrachloride	70

103　[1]Due to insufficient cytotoxicity.
104　[2]Number of substances available for intralaboratory reproducibility analyses.
105
106
107

108　**Table 7-2　　Number of Reference Substances Tested vs Number of Reference**

109　　　　　　　　**Substances Yielding IC$_{50}$ Values in Each GHS Toxicity Category[1] for**

110　　　　　　　　**Two Sets of LD$_{50}$ Values**

GHS Category[1] (LD$_{50}$ in mg/kg)	Initial Oral LD$_{50}$[2]	Reference Oral LD$_{50}$[3]	Results from 3T3 NRU Test Method		Results from NHK NRU Test Method	
			Initial Oral LD$_{50}$[2]	**Reference Oral LD$_{50}$[3]**	**Initial Oral LD$_{50}$[2]**	**Reference Oral LD$_{50}$[3]**
LD$_{50}$ ≤ 5	12	7	12	7	12	7
5 < LD$_{50}$ ≤ 50	12	12	12	12	12	12
50 < LD$_{50}$ ≤ 300	12	12	12	12	12	12
300 < LD$_{50}$ ≤ 2000	12	16	11	15	12	16
2000 < LD$_{50}$ ≤ 5000	12	12	10	10	10	10
LD$_{50}$ > 5000	12	13	11	12	11	12

111　[1]GHS-Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005).
112　[2]Number of reference substances that yielded an IC$_{50}$ value in at least one laboratory based on initial oral LD$_{50}$
113　in **Table 3-2**. Initial oral LD$_{50}$ values, used during the reference substance selection process, were those used by
114　the Registry of Cytotoxicity (RC) (from 1983/84 RTECS®) when applicable. The RC is a database of acute oral
115　LD$_{50}$ values for rats and mice obtained from RTECS® and IC$_{50}$ values from *in vitro* cytotoxicity assays using
116　multiple cell lines and cytotoxicity endpoints for chemicals with known molecular weights (Halle 1998).
117　Values for reference substances not included in the RC came from HSDB or RTECS®.
118　[3]Number of reference substances that yielded an IC$_{50}$ value in at least one laboratory based on reference oral
119　LD$_{50}$ in **Table 4-2**. Reference oral LD$_{50}$ values from rats and mice were derived after evaluating LD$_{50}$ values
120　located through literature searches and references from toxicity databases such as RTECS®.
121

122

123   **7.2   Reproducibility Analyses for the 3T3 and NHK NRU Test Methods**

124

125   Reproducibility of the 3T3 and NHK NRU test methods were performed using ANOVA and

126   CV as described in **Section 5.3.3**. **Table 7-3** reports the results of these analyses for each

127   reference substance and test method.

128

129   7.2.1   ANOVA Results for the 3T3 and NHK NRU Test Methods

130   ANOVA was performed as discussed in Section **5.3.3**. Since the sample sizes from this

131   study were small, usually three observations per laboratory, the ANOVA results may be

132   misleading. There may be some differences that are statistically significant only because

133   there are too few observations within the laboratories to adequately characterize the

134   variability, and/or the within-laboratory variability estimate is small.

135

136   *Differences Among the Laboratories for the 3T3 NRU Test Method*

137   The ANOVA results in **Table 7-3** indicate that there were statistically significant ($p < 0.01$)

138   differences among the laboratories for 26 reference substances. These chemicals are listed in

139   **Table 7-4** along with columns showing the laboratory statistically significantly differing

140   from the other two laboratories (as indicated by the contrast results). Since significant

141   laboratory differences may be produced by insolubility or volatility, **Table 7-4** also indicates

142   whether any laboratory reported insolubility or volatility during conduct of the test.

143   Insolubility was suggested by the presence of precipitates in either the stock solutions or in

144   cell culture. Volatility was identified by the use of plate sealers to contain volatile

145   contamination of lower concentration wells by higher concentrations. Insolubility and

146   volatility were reported for only nine of the 26 chemicals.

147

Table 7-3 Reproducibility Results for the 3T3 and NHK NRU Test Methods

Reference Substance/Laboratory	3T3 NRU Test Method						NHK NRU Test Method					
	Arithmetic Mean IC$_{50}$ (µg/mL)[1]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC$_{50}$ (µg/mL)[1]	ANOVA P[2]	Contrast P[3]	Arithmetic Mean IC$_{50}$ (µg/mL)[1]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC$_{50}$ (µg/mL)[1]	ANOVA P[2]	Contrast P[3]
Acetaminophen	50.1		28	1.70	0.171		526		13	2.72	0.181	
ECBC	40.8	22		1.61		NA	558	15		2.75		NA
FAL	66.2	35		1.82		NA	447	19		2.65		NA
IIVS	43.4	26		1.64		NA	571	14		2.76		NA
Acetonitrile	8484		21	3.93	0.553		10104		8	4.00	0.9641	
ECBC	6433	2		3.81		NA	10868	72		4.04		NA
FAL	9690	58		3.99		NA	10153	19		4.01		NA
IIVS	9330	13		3.97		NA	9290	4		3.97		NA
Acetylsalicylic acid	760		56	2.88	<0.001		613		15	2.79	0.060	
ECBC	646	10		2.81		0.581	631	3		2.80		NA
FAL	1234	24		3.09		<0.001	694	14		2.84		NA
IIVS	401	16		2.60		<0.001	514	15		2.71		NA
5-Aminosalicylic acid	1698		19	3.23	0.054		52.3		47	1.72	0.044	
ECBC	1467	14		3.17		0.092	29.9	22		1.48		0.025
FAL	2070	16		3.32		0.021	78.2	54		1.89		0.033
IIVS	1557	12		3.19		0.312	48.8	16		1.69		0.832
Aminopterin	0.007		54	-2.14	0.036		682		27	2.83	0.0250	
ECBC	0.005	20		-2.28		0.216	889	20		2.95		0.017
FAL	0.012	46		-1.93		0.013	545	8		2.74		0.041
IIVS	0.005	23		-2.33		0.079	611	12		2.79		0.345
Amitriptyline HCl	7.23		14	0.86	0.348		9.76		19	0.99	0.365	
ECBC	6.03	23		0.78		0.163	10.8	31		1.03		NA
FAL	7.86	28		0.90		0.469	7.57	72		0.88		NA
IIVS	7.81	18		0.89		0.445	10.9	10		1.04		NA
Arsenic trioxide	2.51		61	0.40	0.004		10.4		91	1.02	<0.001	
ECBC	2.41	33		0.38		0.527	7.77	33		0.89		0.694
FAL	1.04	7		0.02		0.002	2.55	75		0.41		<0.001
IIVS	4.09	52		0.61		0.006	20.9	31		1.32		0.0006
Atropine sulfate	85.6		49	1.93	0.049		91.9		13	1.96	0.9881	
ECBC	54.1	55		1.73		0.046	85.4	12		1.93		0.8903
FAL	133	31		2.12		0.024	104	85		2.02		0.9069
IIVS	70.0	8		1.85		0.641	83.2	25		1.92		0.9832
Boric acid	2228		69	3.35	0.010		473		8	2.67	0.9306	
ECBC	1497	32		3.18		0.189	440	31		2.64		0.9692
FAL	3987	17		3.60		0.004	517	73		2.71		0.7391
IIVS	1202	48		3.08		0.021	464	2		2.67		0.7680
Busulfan	135		119	2.13	0.002		278		11	2.44	0.659	
ECBC	40.0	48		1.60		0.012	253	27		2.40		NA
FAL	321	56		2.51		< 0.001	268	72		2.43		NA
IIVS	43.7	4		1.64		0.033	313	12		2.50		NA
Cadmium chloride	0.565		39	-0.25	0.124		1.98		10	0.30	0.733	
ECBC	0.480	14		-0.32		NA	2.20	37		0.34		NA

Table 7-3 Reproducibility Results for the 3T3 and NHK NRU Test Methods

Reference Substance/Laboratory	3T3 NRU Test Method						NHK NRU Test Method					
	Arithmetic Mean IC$_{50}$ (μg/mL)[1]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC$_{50}$ (μg/mL)[1]	ANOVA P[2]	Contrast P[3]	Arithmetic Mean IC$_{50}$ (μg/mL)[1]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC$_{50}$ (μg/mL)[1]	ANOVA P[2]	Contrast P[3]
FAL	0.400	32		-0.40		NA	1.88	65		0.27		NA
IIVS	0.817	53		-0.09		NA	1.86	8		0.27		NA
Caffeine	161		18	2.21	0.481		661		21	2.82	0.296	
ECBC	133	10		2.12		NA	817	31		2.91		NA
FAL	157	52		2.20		NA	591	32		2.77		NA
IIVS	191	7.5		2.28		NA	574	1		2.76		NA
Carbamazepine	109		35	2.04	0.049		128		85	2.11	0.432	
ECBC	83.0	14		1.92		NA	66.1	13		1.82		NA
FAL	152	37		2.18		NA	253	129		2.40		NA
IIVS	91.8	12		1.96		NA	63.9	8		1.81		NA
Carbon tetrachloride	NA		NA	NA	NA		NA		NA	NA	NA	
ECBC	NA	NA		NA		NA	NA	NA		NA		NA
FAL	NA	NA		NA		NA	NA	NA		NA		NA
IIVS	NA	NA		NA		NA	NA	NA		NA		NA
Chloral hydrate	187		25	2.27	0.004		137		17	2.14	0.302	
ECBC	151	10		2.18		0.008	140	24		2.15		NA
FAL	241	10		2.38		0.002	159	32		2.20		NA
IIVS	170	12		2.23		0.181	112	2		2.05		NA
Chloramphenicol	161		67	2.21	<0.001		366		13	2.56	0.750	
ECBC	55.3	22		1.74		<0.001	318	45		2.50		NA
FAL	273	30		2.44		0.001	414	44		2.62		NA
IIVS	156	18		2.19		0.165	367	22		2.56		NA
Citric acid	829		41	2.92	0.002		424		25	2.63	0.006	
ECBC	473	29		2.68		0.001	526	16		2.72		0.009
FAL	1148	13		3.06		0.003	312	17		2.49		0.002
IIVS	865	19		2.94		0.298	433	5		2.64		0.483
Colchicine	0.047		85	-1.33	0.001		0.007		22	-2.16	0.174	
ECBC	0.020	11		-1.70		0.0028	0.005	46		-2.28		NA
FAL	0.093	45		-1.03		0.0005	0.008	10		-2.12		NA
IIVS	0.028	1		-1.55		0.0914	0.008	21		-2.09		NA
Cupric sulfate pentahydrate	70.6		85	1.85	<0.001		197		4	2.29	0.374	
ECBC	82.7	4		1.92		0.001	190	10		2.28		NA
FAL	123	44		2.09		<0.001	195	6		2.29		NA
IIVS	5.70	31		0.76		<0.001	207	3		2.32		NA
Cycloheximide	0.293		104	-0.53	0.021		0.082		43	-1.09	0.302	
ECBC	0.125	45		-0.90		0.118	0.053	22		-1.28		NA
FAL	0.647	70		-0.19		0.007	0.120	78		-0.92		NA
IIVS	0.109	23		-0.96		0.076	0.071	19		-1.15		NA
Dibutyl phthalate	78.3		124	1.89	< 0.001		32.6		41	1.51	0.408	
ECBC	23.5	17		1.37		0.012	28.3	27		1.45		NA

Table 7-3 Reproducibility Results for the 3T3 and NHK NRU Test Methods

Reference Substance/Laboratory	3T3 NRU Test Method						NHK NRU Test Method					
	Arithmetic Mean IC_{50} (µg/mL)[1]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (µg/mL)[1]	ANOVA P^2	Contrast P^3	Arithmetic Mean IC_{50} (µg/mL)[1]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (µg/mL)[1]	ANOVA P^2	Contrast P^3
FAL	191	50		2.28		<0.001	47.4	73		1.68		NA
IIVS	20.7	7		1.32		0.005	22.0	6		1.34		NA
Dichlorvos	20.3		57	1.31	0.002		11.1		20	1.05	0.181	
ECBC	9.80	35		0.99		0.001	8.56	27		0.93		NA
FAL	32.8	6		1.52		0.002	12.4	30		1.09		NA
IIVS	18.3	11		1.26		0.823	12.2	3		1.09		NA
Diethyl phthalate	113		28	2.05	0.127		145		44	2.16	0.049	
ECBC	85.5	34		1.93		0.092	174	8		2.24		0.196
FAL	147	26		2.17		0.070	71.5	94		1.85		0.018
IIVS	106	24		2.03		0.846	189	18		2.28		0.127
Digoxin	520		62	2.72	0.043		0.00314		88	-2.50	<0.001	
ECBC	351	39		2.54		0.167	0.00538	13		-2.27		<0.001
FAL	892	36		2.95		0.017	0.00005	36		-4.29		<0.001
IIVS	317	21		2.50		0.144	0.00398	7		-2.40		<0.001
Dimethylformamide	5242		6	3.72	0.296		7856		19	3.90	<0.001	
ECBC	5343	10		3.73		NA	9353	2		3.97		<0.001
FAL	5483	9		3.74		NA	7817	1		3.89		0.508
IIVS	4900	4		3.69		NA	6397	3		3.81		<0.001
Diquat dibromide monohydrate	15.1		120	1.18	0.017		4.73		37	0.67	0.217	
ECBC	3.90	23		0.59		0.040	3.59	23		0.56		NA
FAL	36.1	98		1.56		0.006	6.77	55		0.83		NA
IIVS	5.40	25		0.73		0.190	3.84	8		0.58		NA
Disulfoton	98.6		55	1.99	0.003		378		99	2.58	<0.001	
ECBC	137	55		2.14		NA	140	19		2.15		0.002
FAL	NA	NA		NA		NA	808	26		2.91		<0.001
IIVS	60.4	87		1.78		NA	186	32		2.27		0.018
Endosulfan	8.02		78	0.90	0.046		2.35		43	0.37	0.029	
ECBC	5.30	57		0.72		0.447	3.44	17		0.54		0.020
FAL	15.2	78		1.18		0.018	1.42	50		0.15		0.018
IIVS	3.60	42		0.56		0.080	2.19	20		0.34		0.927
Epinephrine bitartrate	59.4		12	1.77	0.048		90.6		24	1.96	0.119	
ECBC	51.5	12		1.71		0.018	115	9		2.06		NA
FAL	63.4	11		1.80		0.165	81.7	35		1.91		NA
IIVS	63.4	3		1.80		0.149	75.0	16		1.88		NA
Ethanol	6731		23	3.83	0.075		10184		18	4.01	0.035	
ECBC	5360	33		3.73		NA	8290	5		3.92		0.019
FAL	8420	14		3.93		NA	12013	19		4.08		0.029
IIVS	6413	5		3.81		NA	10250	9		4.01		0.752
Ethylene glycol	25292		26	4.40	0.007		42600		15	4.63	0.063	
ECBC	18325	9		4.26		0.004	38000	12		4.58		NA
FAL	31650	24		4.50		0.010	49800	9		4.70		NA

Table 7-3 Reproducibility Results for the 3T3 and NHK NRU Test Methods

Reference Substance/Laboratory	3T3 NRU Test Method						NHK NRU Test Method					
	Arithmetic Mean IC$_{50}$ (µg/mL)[1]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC$_{50}$ (µg/mL)[1]	ANOVA P[2]	Contrast P[3]	Arithmetic Mean IC$_{50}$ (µg/mL)[1]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC$_{50}$ (µg/mL)[1]	ANOVA P[2]	Contrast P[3]
IIVS	25900	12		4.41		0.505	40000	13		4.60		NA
Fenpropathrin	27.2		49	1.43	0.301		2.60		39	0.41	0.031	
ECBC	22.6	11		1.35		NA	3.73	27		0.57		0.013
FAL	42.4	63		1.63		NA	2.23	28		0.35		0.375
IIVS	16.7	12		1.22		NA	1.82	17		0.26		0.044
Gibberellic Acid	7842		3	3.89	0.621		2866		2	3.46	0.862	
ECBC	8027	11		3.90		NA	2850	14		3.45		NA
FAL	NA	NA		NA		NA	2940	9		3.47		NA
IIVS	7657	10		3.88		NA	2807	4		3.45		NA
Glutethimide	192		43	2.28	< 0.001		177		5	2.25	0.968	
ECBC	167	4		2.22		0.029	187	34		2.27		NA
FAL	284.3	7		2.45		<0.001	170	14		2.23		NA
IIVS	125.3	7		2.10		<0.001	176	16		2.24		NA
Glycerol	28904		33	4.46	0.846		27108		31	4.43	0.200	
ECBC	20000	15		4.30		NA	34267	45		4.53		NA
FAL	38878	73		4.59		NA	18023	46		4.26		NA
IIVS	27833	39		4.44		NA	29033	16		4.46		NA
Haloperidol	6.26		24	0.80	0.006		3.57		7	0.55	0.935	
ECBC	5.30	12		0.72		0.030	3.69	27		0.57		NA
FAL	8.00	8		0.90		0.002	3.72	49		0.57		NA
IIVS	5.50	12		0.74		0.061	3.29	35		0.52		NA
Hexachlorophene	4.48		27	0.65	0.174		0.031		41	-1.50	0.097	
ECBC	5.00	48		0.70		NA	0.027	16		-1.57		NA
FAL	5.30	33		0.72		NA	0.046	44		-1.34		NA
IIVS	3.10	9		0.49		NA	0.021	11		-1.67		NA
Lactic acid	3073		12	3.49	0.160		1308		1	3.12	0.904	
ECBC	2943	11		3.47		NA	1290	4		3.11		NA
FAL	3487	16		3.54		NA	1320	5		3.12		NA
IIVS	2790	9		3.45		NA	1313	11		3.12		NA
Lindane	161		58	2.21	0.066		19.3		20	1.29	0.203	
ECBC	125	95		2.10		NA	19.1	17		1.28		NA
FAL	266	36		2.43		NA	23.2	31		1.37		NA
IIVS	90.4	122		1.96		NA	15.6	15		1.19		NA
Lithium carbonate	NA		NA	NA	NA	NA	477		13	2.68	0.295	
ECBC	564	12		2.75		NA	411	29		2.61		NA
FAL	NA	NA		NA		NA	486	20		2.69		NA
IIVS	NA	NA		NA		NA	535	6		2.73		NA
Meprobamate	539		54	2.73	<0.001		516		61	2.71	0.027	
ECBC	353	14		2.55		0.001	761	15		2.88		0.0758
FAL	877	15		2.94		<0.001	163	116		2.21		0.0098
IIVS	386	2		2.59		0.005	624	14		2.80		0.1648

Table 7-3 Reproducibility Results for the 3T3 and NHK NRU Test Methods

Reference Substance/Laboratory	3T3 NRU Test Method						NHK NRU Test Method					
	Arithmetic Mean IC$_{50}$ (µg/mL)[1]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC$_{50}$ (µg/mL)[1]	ANOVA P[2]	Contrast P[3]	Arithmetic Mean IC$_{50}$ (µg/mL)[1]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC$_{50}$ (µg/mL)[1]	ANOVA P[2]	Contrast P[3]
Mercury chloride	4.32		33	0.64	0.021		5.87		15	0.77	0.120	
ECBC	3.50	5		0.54		0.083	6.87	15		0.84		NA
FAL	6.00	31		0.78		0.008	5.40	19		0.73		NA
IIVS	3.50	3		0.54		0.110	5.35	2		0.73		NA
Methanol	NA		NA	NA	NA	NA	1616		42	3.21	0.007	
ECBC	NA	NA		NA		NA	NA	NA		NA		NA
FAL	NA			NA		NA	1133	19		3.05		NA
IIVS	NA			NA		NA	2100	11		3.32		NA
Nicotine	378		25	2.58	0.128		113		17	2.05	0.700	
ECBC	272	24		2.43		NA	94.3	26		1.97		NA
FAL	412	33		2.61		NA	134	59		2.13		NA
IIVS	450	12		2.65		NA	112	25		2.05		NA
Paraquat	23.3		8	1.37	1.000		66.1		40	1.82	0.047	
ECBC	21.3	34		1.33		NA	48.3	13		1.68		0.089
FAL	24.9	67		1.40		NA	96.6	39		1.98		0.018
IIVS	23.7	64		1.37		NA	53.4	10		1.73		0.279
Parathion	61.8		111	1.79	0.014		31.4		8	1.50	0.845	
ECBC	22.7	53		1.36		0.064	34.0	30		1.53		NA
FAL	141	70		2.15		0.005	31.2	38		1.49		NA
IIVS	22	22		1.34		0.081	29.0	29		1.46		NA
Phenobarbital	612		21	2.79	0.232		478		39	2.68	0.027	
ECBC	634	21		2.80		NA	693	26		2.84		0.010
FAL	726	35		2.86		NA	360	27		2.56		0.072
IIVS	476	23		2.68		NA	381	18		2.58		0.173
Phenol	70.9		41		0.011		77.7		22	1.89	0.094	
ECBC	50.2	22		1.70		0.022	59.1	36		1.77		NA
FAL	104	24		2.02		0.004	93.2	6		1.97		NA
IIVS	58.1	12		1.76		0.206	80.8	6		1.91		NA
Phenylthiourea	119		90	2.08	0.007		346		19	2.54	0.133	
ECBC	30.1	66		1.48		0.004	363	16		2.56		NA
FAL	239	28		2.38		0.006	401	21		2.60		NA
IIVS	89	25		1.95		0.718	272	26		2.44		NA
Physostigmine	28.8		30	1.46	0.149		172		22	2.24	0.623	
ECBC	28.2	53		1.45		NA	164	3		2.21		NA
FAL	37.8	5		1.58		NA	213	112		2.33		NA
IIVS	20.4	33		1.31		NA	139	6		2.14		NA
Potassium chloride	3635		7	3.56	0.846		2279		13	3.36	0.396	
ECBC	3352	14		3.53		NA	2560	17		3.41		NA
FAL	3842	31		3.58		NA	2287	28		3.36		NA
IIVS	3710	11		3.57		NA	1990	8		3.30		NA
Potassium cyanide	64.3		127	1.81	<0.001		45.1		86	1.65	0.340	
ECBC	15.3	25		1.18		0.001	29.3	24		1.47		NA

Table 7-3 Reproducibility Results for the 3T3 and NHK NRU Test Methods

Reference Substance/Laboratory	3T3 NRU Test Method						NHK NRU Test Method					
	Arithmetic Mean IC$_{50}$ (µg/mL)[1]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC$_{50}$ (µg/mL)[1]	ANOVA P[2]	Contrast P[3]	Arithmetic Mean IC$_{50}$ (µg/mL)[1]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC$_{50}$ (µg/mL)[1]	ANOVA P[2]	Contrast P[3]
FAL	159	52		2.20		<0.001	89.0	112		1.95		NA
IIVS	18.9	5		1.28		0.006	16.9	13		1.23		NA
Procainamide HCl	443		11	2.65	0.007		1764		16	3.25	0.053	
ECBC	400	4		2.60		0.008	1480	14		3.17		NA
FAL	431	1		2.63		0.396	1787	12		3.25		NA
IIVS	497	8		2.70		0.003	2027	11		3.31		NA
2-Propanol	3563		23	3.55	0.001		5541		26	3.74	0.033	
ECBC	2610	9		3.42		< 0.001	5263	11		3.72		0.797
FAL	3970	4		3.60		0.004	4273	27		3.63		0.026
IIVS	4110	4		3.61		0.002	7087	7		3.85		0.018
Propranolol HCl	14.9		16	1.17	0.488		36.9		21	1.57	0.003	
ECBC	13.6	32		1.13		NA	38.27	12		1.58		0.325
FAL	13.5	51		1.13		NA	43.8	6		1.64		0.006
IIVS	17.6	21		1.25		NA	28.6	11		1.46		0.001
Propylparaben	29.9		64	1.48	0.001		16.8		16	1.23	0.066	
ECBC	20.9	16		1.32		0.045	18.1	13		1.26		NA
FAL	51.8	29		1.71		< 0.001	18.6	15		1.27		NA
IIVS	17.1	12		1.23		0.003	13.8	9		1.14		NA
Sodium arsenite	0.873		55	-0.06	0.028		0.532		44	-0.27	0.061	
ECBC	0.500	6		-0.30		0.032	0.790	32		-0.10		NA
FAL	1.40	57		0.15		0.012	0.336	56		-0.47		NA
IIVS	0.700	17		-0.15		0.478	0.470	14		-0.33		NA
Sodium chloride	4764		3	3.68	0.759		2724		51	3.44	0.045	
ECBC	4790	5		3.68		NA	3583	7		3.55		0.141
FAL	4625	13		3.67		NA	1118	124		3.05		0.017
IIVS	4877	9		3.69		NA	3470	9		3.54		0.161
Sodium dichromate dihydrate	0.602		9	-0.22	0.822		0.737		19	-0.13	0.258	
ECBC	0.603	14		-0.22		NA	0.784	14		-0.11		NA
FAL	0.657	37		-0.18		NA	0.851	36		-0.07		NA
IIVS	0.547	17		-0.26		NA	0.576	17		-0.24		NA
Sodium fluoride	79.8		22	1.90	0.016		47.4		15	1.68	0.313	
ECBC	61.3	9		1.79		0.007	48.7	14		1.69		NA
FAL	96.1	18		1.98		0.019	39.7	24		1.60		NA
IIVS	82.0	7		1.91		0.463	53.7	13		1.73		NA
Sodium hypochlorite	1211		57	3.08	0.040		1580		20	3.20	0.313	
ECBC	823	13		2.92		0.257	1863	31		3.27		NA
FAL	805	46		2.91		0.119	1243	46		3.09		NA
IIVS	2005	44		3.30		0.015	1633	11		3.21		NA
Sodium oxalate	40.8		23	1.61	0.643		355		1	2.55	0.926	
ECBC	42.0	41		1.62		NA	355	15		2.55		NA

Table 7-3 Reproducibility Results for the 3T3 and NHK NRU Test Methods

Reference Substance/Laboratory	3T3 NRU Test Method						NHK NRU Test Method					
	Arithmetic Mean IC_{50} $(\mu g/mL)$[1]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} $(\mu g/mL)$[1]	ANOVA P[2]	Contrast P[3]	Arithmetic Mean IC_{50} $(\mu g/mL)$[1]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} $(\mu g/mL)$[1]	ANOVA P[2]	Contrast P[3]
FAL	31.0	28		1.49		NA	350	42		2.54		NA
IIVS	49.5	53		1.69		NA	360	26		2.56		NA
Sodium selenate	34.5		60	1.54	<0.001		11.2		40	1.05	0.134	
ECBC	12.7	13		1.10		<0.001	7.47	12		0.87		NA
FAL	54.2	19		1.73		< 0.001	16.1	59		1.21		NA
IIVS	36.5	14		1.56		0.026	10.0	13		1.00		NA
Strychnine	199		83	2.30	<0.001		69.3		39	1.84	0.364	
ECBC	389	21		2.59		<0.001	100	76		2.00		NA
FAL	124	16		2.09		0.018	52.5	53		1.72		NA
IIVS	83.5	6		1.92		<0.001	55.1	6		1.74		NA
Thallium Sulfate	7.50		72	0.88	0.165		0.16		23	-0.80	0.405	
ECBC	2.80	24		0.45		NA	0.198	51		-0.70		NA
FAL	13.4	78		1.13		NA	0.153	20		-0.82		NA
IIVS	6.30	28		0.80		NA	0.127	16		-0.90		NA
Trichloroacetic acid	928		27	2.97	0.005		427		24	2.63	0.134	
ECBC	762	13		2.88		0.022	348	18		2.54		NA
FAL	1220	6		3.09		0.002	541	28		2.73		NA
IIVS	801	14		2.90		0.069	394	13		2.60		NA
1,1,1-Trichloroethane	15538		52	4.19	<0.001		NA		NA	NA	NA	
ECBC	NA	NA		NA		NA	8137	7		3.91		NA
FAL	21250	11		4.33		NA	NA	NA		NA		NA
IIVS	9827	2		3.99		NA	NA	NA		NA		NA
Triethylenemelamine	0.568		135	-0.25	<0.001		1.95		12	0.29	0.562	
ECBC	0.086	11		-1.07		<0.001	1.69	57		0.23		NA
FAL	1.45	18		0.16		<0.001	2.03	23		0.31		NA
IIVS	0.169	29		-0.77		0.002	2.13	23		0.33		NA
Triphenyltin hydroxide	0.022		29	-1.66	0.688		0.013		55	-1.89	0.088	
ECBC	0.026	17		-1.59		NA	0.021	32		-1.68		NA
FAL	0.026	81		-1.59		NA	0.007	106		-2.15		NA
IIVS	0.015	55		-1.83		NA	0.011	32		-1.96		NA
Valproic acid	1177		76	3.07	< 0.001		533		28	2.73	0.081	
ECBC	547	12		2.74		NA	468	25		2.67		0.331
FAL	1807	10		3.26		NA	702	23		2.85		0.032
IIVS	NA	NA		NA		NA	430	17		2.63		0.135
Verapamil HCl	35.2		10	1.55	0.230		68.7		14	1.84	0.624	
ECBC	32.0	18		1.51		NA	60.5	22		1.78		NA
FAL	34.6	5		1.54		NA	79.4	42		1.90		NA
IIVS	38.9	11		1.59		NA	66.2	8		1.82		NA
Xylene	NA		NA	NA	NA	NA	NA		NA	NA	NA	
ECBC	NA	NA		NA		NA	NA	NA		NA		NA
FAL	NA	NA		NA		NA	NA	NA		NA		NA
IIVS	724	12		2.86		NA	486	38		2.69		NA

148   [1]Results reported on the same row with chemical names are the means of all the laboratories. Results
149   reported on the same row as laboratories are the laboratory means.
150   [2]$p < 0.01$ indicated statistical significance.
151   [3]Contrasts were performed if ANOVA was significant ($p < 0.01$) to determine which laboratory was
152   different from the other two laboratories. Significant contrasts were denoted by $p < 0.01$. If only two
153   laboratories reported results, no contrast tests were necessary.
154   Abbreviations: Laboratories: ECBC- U.S. Army Edgewood Chemical Biological Center; FAL – FRAME
155   Alternatives Laboratory; IIVS – Institute for In Vitro Sciences. NA - no acceptable IC_{50} results reported or
156   calculation was not performed (e.g., for contrast results).
157

158   **Table 7-4   Reference Substances with Significant Differences between Laboratories**

159   **for 3T3 NRU Test Method Results**

Reference Substance	Significant Contrast Results[1]			Insoluble/ Volatile[2]
	ECBC	FAL	IIVS	
Acetylsalicylic acid		H	L	
Arsenic trioxide		L	H	Precipitate
Busulfan		H		
Chloral hydrate	L	H		
Chloramphenicol	L	H		
Citric acid	L	H		
Colchicine	L	H		
Cupric sulfate pentahydrate	X	H	L	
Dibutyl phthalate		H	L	Precipitate
Dichlorvos	L	H		Precipitate
Disulfoton[3]				Precipitate
Ethylene glycol	L			
Glutethimide		H	L	
Haloperidol		H		
Meprobamate	L	H	X	
Phenylthiourea	L	H		
Potassium cyanide	L	H	X	Precipitate /Volatility
Procainamide HCl	L		H	
2-Propanol	L	X	H	Volatility
Propylparaben		H	L	
Sodium selenate	L	H		
Strychnine	H		L	Precipitate
Trichloroacetic acid		H		
1,1,1-Trichloroethane[4]				Precipitate
Triethylenemelamine	L	H		
Valproic acid[5]				Precipitate

160   [1]Laboratories significantly different from the other two at p < 0.01. H – Laboratory reported the highest mean
161   IC$_{50}$. L – Laboratory reported the lowest mean IC$_{50}$. X – Laboratory reported a mean IC$_{50}$ between the values
162   of the other two laboratories.
163   [2]From **Table 5-8**. Precipitate reported by at least one laboratory is indicated by "Precipitate". Use of plate
164   sealers by at least one laboratory to prevent volatile contamination of control wells indicated by "Volatility".
165   [3]Significant ANOVA (p < 0.01), but no contrast analysis since only two laboratories (ECBC and IIVS) reported
166   IC$_{50}$ values.
167   [4]Significant ANOVA (p < 0.01), but no contrast results since only two laboratories (FAL and IIVS) reported
168   IC$_{50}$ values.
169   [5]Significant ANOVA (p < 0.01), but no contrast results since only two laboratories (ECBC and FAL) reported
170   IC$_{50}$ values.
171   Laboratories: ECBC- U.S. Army Edgewood Chemical Biological Center; FAL – FRAME Alternatives
172   Laboratory; IIVS – Institute for In Vitro Sciences.
173
174
175   For the 26 substances that yielded significantly different results among the laboratories,

176   contrast analyses indicated that ECBC and FAL were frequently different from the other

177   laboratories. ECBC tended to report the lowest IC$_{50}$ values among the laboratories while

178   FAL tended to report the highest values of the three laboratories. ECBC reported

179   significantly different results from the other two laboratories for 15 of the 26 substances.

180   For 13 of the 15 substances, ECBC reported the lowest mean IC_{50} value among the three

181   laboratories. FAL reported significantly different results from the other two laboratories for

182   20 of the 26 substances. For 18 of the 20 substances, FAL reported the highest mean IC_{50}

183   value among the three laboratories. IIVS reported significantly different results for 11 of

184   the 26 substances, with no great majority of highest or lowest IC_{50} values.

185

186   *Differences Among the Laboratories for the NHK NRU Test Method*

187   The ANOVA results in **Table 7-3** indicate that there were statistically significant ($p < 0.01$)

188   laboratory differences for seven substances. These substances are listed in **Table 7-5** along

189   with columns showing the laboratory statistically significantly differing from the other two

190   laboratories (as indicated by the contrast results), and indications of whether any laboratory

191   reported insolubility or volatility during conduct of the assay. Insolubility was reported for

192   three of the seven substances.

193
194

195   **Table 7-5   Reference Substances with Significant Differences between Laboratories**

196   **   for NHK NRU Test Method Results**

Reference Substance	Significant Contrast Results[1]			Solubility/ Volatility[2]
	ECBC	**FAL**	**IIVS**	
Arsenic trioxide		L	H	Precipitate
Citric acid	H	L		Precipitate
Digoxin	H	L		
Dimethylformamide	H		L	
Disulfoton	L	H		Precipitate
Methanol[3]				
Propranolol HCl		H	L	

197 [1]Laboratories significantly different from the other two at $p < 0.01$. H – Laboratory reported the highest mean
198 IC_{50}. L – Laboratory reported the lowest mean IC_{50}. X – Laboratory reported a mean IC_{50} between the values
199 of the other two laboratories.
200 [2]From **Table 5-8**. Precipitate reported by at least one laboratory is indicated by "Precipitate". Use of plate
201 sealers by at least one laboratory to prevent volatile contamination of control wells indicated by "Volatility".
202 [3]Significant ANOVA ($p < 0.01$), but no contrast results since only two laboratories (FAL and IIVS) reported
203 IC_{50} values.
204 Laboratories: ECBC – U.S. Army Edgewood Chemical Biological Center; FAL – FRAME Alternatives
205 Laboratory; IIVS – Institute for In Vitro Sciences.
206

207 For the seven substances that yielded significantly different results among the laboratories,

208 ECBC and FAL were frequently different from the other laboratories. ECBC tended to

209 report the highest IC_{50} value among the laboratories (4/7 substances) while FAL tended to

210 report the lowest values among the three laboratories (3/7 substances).

211

212 7.2.2   CV Results for the 3T3 and NHK NRU Test Methods

213 CV was calculated as described in Section **5.3.3**. **Table 7-3** provides the intra- and inter-

214 laboratory CV values for individual substances. **Table 7-6** summarizes the CV results for

215 each test method. **Table 7-6** shows that median and mean CV values were often similar.

216 Median CV values appeared always lower than the corresponding means, which indicated

217 that large individual CV values skewed the CV distributions somewhat to the right.

218

219 *Intralaboratory CV*

220 **Table 7-6** shows that both test methods had similar ranges for the intralaboratory CV. The

221 mean intralaboratory CV values were the same, 26%. The median intralaboratory CVs were

222 also similar: 23% for the 3T3 NRU test method and 24% for the NHK NRU test method. Of

223 the three laboratories, FAL had the highest mean and median CV for both test methods and

224 IIVS had the lowest mean and median CV for both test methods.

225

226 *Interlaboratory CV*

227 The mean and median interlaboratory CV for the reference substances was lower for the

228 NHK NRU test method (mean = 28%; median = 21%) than for the 3T3 NRU test method

229 (mean = 46%; median = 40%) (see **Table 7-6**).

230

230   **Table 7-6   Summary of CV Results for the 3T3 and NHK NRU Test Methods**

CV	3T3 NRU Test Method				NHK NRU Test Method			
	N	Mean	Median	Range	N	Mean	Median	Range
Intralaboratory CV	202	26%	23%	1-122%	208	26%	24%	1-129%
ECBC	68	23%	17%	2-95%	69	23%	19%	2-76%
FAL	66	33%	30%	1-98%	69	42%	32%	1-129%
IIVS	68	21%	13%	1-122%	70	14%	13%	1-38%
Interlaboratory CV	68	46%	40%	2-135%	68	28%	21%	1-99%

231   Abbreviations: N- number of values. Laboratories: ECBC- U.S. Army Edgewood Chemical Biological Center;
232   FAL – FRAME Alternatives Laboratory; IIVS – Institute for In Vitro Sciences.
233   Note: For the 3T3 NRU test method, the following laboratories/substances did not obtain sufficient IC_{50} data for
234   the calculation of an intralaboratory CV: carbon tetrachloride at any laboratory; disulfoton at FAL; gibberellic
235   acid at FAL; lithium carbonate at FAL and IIVS; methanol at any laboratory; 1,1,1-trichloroethane at ECBC;
236   valproic acid at IIVS; and xylene at ECBC and FAL. For the NHK assay, the following laboratories/substances
237   did not obtain sufficient IC_{50} data for the calculation of an intralaboratory CV: carbon tetrachloride at any
238   laboratory; methanol at ECBC; 1,1,1-trichloroethane at FAL and IIVS; and xylene at ECBC and FAL. For the
239   3T3 NRU test method, the following substances did not obtain sufficient IC_{50} data for the calculation of an
240   interlaboratory CV: carbon tetrachloride, lithium carbonate; methanol; and xylene. For the NHK assay, the
241   following substances did not yield sufficient IC_{50} data for the calculation of an interlaboratory CV: carbon
242   tetrachloride; 1,1,1-trichloroethane; and xylene.
243
244

245   *Variation of CV with Chemical Property*

246   To identify the chemical characteristics that may yield high or low CV values, CV values

247   were analyzed to determine their association with the following chemical attributes: physical

248   state (i.e., solid or liquid), solubility, volatility, chemical class, GHS acute oral toxicity class

249   (UN 2005), molecular weight, log Kow, IC_{50}, and boiling point. For categorical

250   characteristics such as physical form, solubility (i.e., precipitate/no precipitate), volatile/not

251   volatile, and chemical class, the mean CV values and CV ranges for the groups were

252   compared to one another and to the overall mean CV and CV range for each test method. No

253   statistical analyses were performed. For chemical characteristics measured by continuous

254   variables, such as molecular weight, log K_{ow}, and IC_{50}, and boiling point, Spearman

255   correlation analyses were performed.

256

257   *Results of Intralaboratory CV Analysis*

258   **Table 7-7** shows the analysis of intralaboratory CV. The analysis of intralaboratory CV uses

259   one mean intralaboratory CV for each reference substances that was calculated from the

260   intralaboratory CV values from each laboratory. With the exception of the amides, which

261   had relatively low intralaboratory CV values (for both 3T3 and NHK NRU test methods),

262   and organophosphates and halogenated hydrocarbons (for the 3T3 NRU test method only),

263   which had relatively high intralaboratory CV values, there seemed to be little difference in

264   CV values for the categorical physical/chemical/toxicological attributes. The mean

265   intralaboratory CV values for solids and liquids were similar (26 vs. 24% for the 3T3 NRU

266   test method; 27 vs. 23% for the NHK NRU test method). The mean intralaboratory CV

267   values for reference substances for which precipitates were observed were similar to the

268   mean intralaboratory CV values for substances for which no precipitates were observed (29

269   vs. 23% for the 3T3 NRU test method; 24 vs. 27% for the NHK NRU test method). The

270   mean intralaboratory CV values for substances that exhibited volatility (i.e., indicated by

271   laboratory use of film plate sealers to prevent contamination of control wells) were relatively

272   similar to those that did not (31 vs. 24% for the 3T3 NRU test method; 27 vs. 26% for the

273   NHK NRU test method). Similarly, the substances grouped by GHS toxicity category (UN

274   2005) had mean intralaboratory CV values that were similar (19-33% for the 3T3 NRU test

275   method; 18-31% for the NHK NRU test method) to the overall mean CV values (26% for

276   both the 3T3 and NHK NRU test methods).

277

278   Reference substances in the amide chemical class had unusually low mean intralaboratory

279   CV values for both the 3T3 NRU test method (13%) and NHK NRU test method (10%)

280   compared with the overall mean CV (26% for both test methods), but there were only three

281   substances in the class (acetaminophen, dimethylformamide, and procainamide HCl).

282   Reference substances in the organophosphate chemical class had unusually high mean

283   intralaboratory CV values for the 3T3 NRU test method (46%), but not for the NHK NRU

284   test method (26%) compared with the overall mean CV (26% for the 3T3 and NHK NRU test

285   methods). There were only three substances in the class (dichlorvos, disulfoton, and

286   parathion), but two of the three substances had relatively high mean intralaboratory CV

287   values (17, 48 and 71%). Halogenated hydrocarbons also had high mean intralaboratory CV

288   for the 3T3 NRU test method (46%), but not for the NHK NRU test method (14%) compared

289   with the overall mean intralaboratory CV for each test method (26%). However, the mean

290   intralaboratory CV for the 3T3 NRU test method was calculated from only two values; 7%

291   for 1,1,1-trichloroethane and 84% for lindane. No laboratory obtained sufficient toxicity for

292   the calculation of an IC_{50} for the carbon tetrachloride, the third halogenated hydrocarbon.

293

294   **Table 7-7   Intralaboratory CV by Chemical Characteristics for the 3T3 and NHK**

295   **NRU Test Methods**

Class/Attribute	3T3 NRU Test Method			NHK NRU Test Method		
	N^a	Range	Mean	N^b	Range	Mean
All chemicals	70	1-122%	26%	71	1-129%	26%
Chemical form						
Solid	53	4-84	26	53	6-50	27
Liquid	17	6-71	24	18	2-40	23
Solubility						
Precipitate[c]	24	7-84	29	2[a]	2-47	24
No precipitate	46	4-55	23	50	7-57	27
Volatility[d]						
Volatile	10	6-84	31	9	11-50	27
Nonvolatile	62	4-71	24	63[b]	2-57	26
Chemical Class						
Alcohols	9	6-42	22	10	10-37	21
Carboxylic acids	12	10-41	20	12	7-48	26
Heterocyclics	14	6-59	30	14	13-50	31
Organophosphorous	3	17-71	46	3	20-32	26
Amides	3	4-28	13	3	2-16	10
Halogenated hydrocarbons	2	7-84	46	2	7-21	14
Inorganics	15	9-43	24	15	6-50	29
Toxicity Class						
≤ 5 mg/kg	7	9-71	33	7	20-40	30
$> 5 - \leq 50$	12	13-59	32	12	12-50	31
$> 50 - \leq 300$	12	11-84	33	12	17-37	25
$> 300 - \leq 2000$	16	4-51	21	16	6-57	25
$> 2000 - \leq 5000$	10[a]	9-32	19	10[a]	7-50	31
> 5000	13[b]	6-42	19	14	2-40	18
Correlations	N	r_s	P value	N	r_s	P value
Molecular weight	70[a,b]	0.323	0.006	71[b]	0.199	0.097
Log K_{ow}	50[e]	0.117	0.421	51[e]	0.311	0.026
IC_{50}	70[a,b]	-0.436	0.0002	71[b]	-0.362	0.002
Boiling point	27	0.576	0.002	28	0.277	0.154

296   [a]One intralaboratory CV for each chemical was calculated by averaging the CV values for the laboratories that
297   reported sufficient data for the calculation of a CV. No CV was calculable for carbon tetrachloride or methanol.
298   [b]One intralaboratory CV for each chemical was calculated by averaging the CV values for the laboratories that
299   reported sufficient data for the calculation of a CV. No CV was calculable for carbon tetrachloride.
300   [c]Denoted by laboratory reports of precipitate in the stock reference substance solutions or in cell culture (see
301   **Table 5-8**).
302   [d]Denoted by laboratory reports of using plate sealers to avoid contamination of the VC wells (see **Table 5-8**).
303   [e]Number of reference substances with CV values and log K_{ow} data.
304   [f]Number of reference substances with CV values and boiling point data.
305

306   For the characteristics amenable to correlation analysis, none of the correlation coefficients

307   were large (absolute value of r_s < 0.6), but several were statistically significantly different

308   from zero for the 3T3 NRU test method. Molecular weight (p = 0.006), IC_{50} (p = 0.0002),

309   and boiling point (p = 0.002) exhibited statistically significant correlations (p < 0.05) to

310   intralaboratory CV for the 3T3 NRU test method. For molecular weight, the higher

311   molecular weight substances had higher intralaboratory CV values. For IC_{50}, however, the

312   substances with lower IC_{50} values had higher CV values. The inverse correlation between

313   intralaboratory CV values and IC_{50} is consistent with the common observation that

314   measurements with very low values tend to have high CV values. The fact that substances

315   with higher boiling points had higher CV values was consistent with the categorical analysis

316   of volatility. The substances that exhibited volatile characteristics (i.e., high reference

317   substance concentration wells contaminated the VC wells) in the 3T3 NRU test method had

318   higher mean intralaboratory CV values (31%) than the substances that did not exhibit volatile

319   characteristics (24%), but the difference did not seem large.

320

321   Likewise, for the NHK NRU test method, two of the characteristics amenable to correlation

322   analysis were statistically significantly different from zero, but the correlation coefficients

323   did not have large magnitudes (absolute value of r_s < 0.4). Log K_{ow} (p = 0.026) and IC_{50}

324   (p = 0.002) exhibited statistically significant correlations (p < 0.05) to intralaboratory CV for

325   the NHK NRU test method. Log K_{ow} was positively correlated to the mean intralaboratory

326   CV for each substance, but IC_{50}, was negatively correlated to the mean IC_{50} for each

327   substance.

328

329   *Results of Interlaboratory CV Analysis*

330   **Table 7-8** shows the analysis of interlaboratory CV. With the exception of chemical class,

331   there seemed to be little difference in interlaboratory CV values for most of the categorical

332   physical/chemical characteristics. The mean interlaboratory CV values for solids and liquids

333   were similar (48 vs. 46% for the 3T3 NRU test method and 28 vs. 27% for the NHK NRU

334   test method). The mean interlaboratory CV values for substances for which precipitates were

335   observed was similar to the mean interlaboratory CV values for substances for which no

336   precipitates were observed (56 vs. 43% for the 3T3 NRU test method and 29 vs. 28% for the

337   NHK NRU test method). The mean interlaboratory CV values for substances that exhibited

338   volatile characteristics appeared similar to those that did not (51 vs. 46% for the 3T3 NRU

339   test method and 32 vs. 28% for the NHK NRU test method).

340

341   Reference substances in the amide chemical class had unusually low mean interlaboratory

342   CV values for both the 3T3 NRU test method (15%) and NHK NRU test method (16%)

343   compared with the overall mean interlaboratory CV (46% for the 3T3 NRU test method and

344   28% for the NHK NRU test method). Chemicals in the organophosphate chemical class had

345   unusually high mean interlaboratory CV values for the 3T3 NRU test method (74%) and

346   moderately higher mean interlaboratory CV values for the NHK NRU test method (42%)

347   compared with the overall mean interlaboratory CV (46% for the 3T3 NRU test method and

348   28% for the NHK NRU test method). The high mean interlaboratory CV value for

349   organophosphates in the NHK NRU test method, however, was produced largely by the high

350   interlaboratory CV of 99% for disulfoton. The interlaboratory CV values for dichlorvos and

351   parathion were 20% and 8%, respectively. Heterocyclic compounds also had higher mean

352   interlaboratory CV values for the 3T3 NRU test method but not for the NHK NRU test

353   method. As a group, the 14 heterocyclic compounds had a mean interlaboratory CV of 61%

354   while the overall mean interlaboratory CV for the 3T3 NRU test method was 46%. Although

355   there were a few low CV values (e.g., 8, 18) in the heterocyclic group, there were seven

356   values greater than the overall mean CV of 46%. The median interlaboratory CV for the

357   heterocyclic group was 52%.

358

358 **Table 7-8   Interlaboratory CV by Chemical Characteristics for the 3T3 and NHK**

359 **NRU Test Methods**

Class/Attribute	3T3 NRU Test Method			NHK NRU Test Method		
	N	Range	Mean	N	Range	Mean
All chemicals	68[a]	2-135%	46%	69[b]	1-99%	28%
Chemical Form						
Solids	52	3-135	48	53	1-91	28
Liquids	16	6-124	46	16	1-99	27
Solubility						
Precipitate[c]	22	3-127	56	19	1-99	29
No precipitate	47	3-135	43	50	1-88	28
Volatility						
Volatile[d]	10	21-127	51	9	8-86	32
Nonvolatile	58	3-135	46	60	1-99	28
Chemical Class						
Alcohols	9	12-119	38	10	11-42	22
Carboxylic acids	12	12-124	46	12	1-61	27
Heterocyclics	14	8-135	61	14	5-85	32
Organophosphorous	3	57-111	74	3	8-99	42
Amides	3	6-28	15	3	13-19	16
Halogenated hydrocarbons	2	52-58	55	1	20	20
Inorganics	14	3-127	48	15	4-91	29
Toxicity Class						
$\leq$ 5 mg/kg	7	12-135	69	7	12-99	37
> 5 - $\leq$ 50	12	33-127	78	12	8-91	41
> 50 - $\leq$ 300	12	8-120	37	12	10-41	26
> 300 - $\leq$ 2000	15	11-85	38	15	1-61	20
> 2000 - $\leq$ 5000	9	3-69	29	9	1-85	27
> 5000	13	3-124	39	13	2-44	25
Correlations		r_s	P value		r_s	P value
Molecular weight	68	0.193	0.115	69	0.136	0.265
Log K_{ow}	49[e]	0.194	0.182	49	0.170	0.244
IC_{50}	68	-0.295	0.015	69	-0.271	0.024
Boiling point	24[f]	0.467	0.021	26	-0.131	0.525

360 [a]The following chemicals did not have sufficient IC_{50} data for the calculation of an interlaboratory CV: carbon
361 tetrachloride, lithium carbonate; methanol; and xylene.
362 [b]The following substances did not yield sufficient IC_{50} data for the calculation of an interlaboratory CV: carbon
363 tetrachloride; 1,1,1-trichloroethane; and xylene.
364 [c]Denoted by laboratory reports of precipitate in the stock reference substance solutions or in cell culture (see
365 **Table 5-8**).
366 [d]Denoted by laboratory reports of using plate sealers to avoid contamination of the VC wells (see **Table 5-8**).
367 [e]Number of reference substances with CV values and log K_{ow} data.
368 [f]Number of reference substances with CV values and boiling point data.

369

370 Mean interlaboratory CV values tended to be large for chemicals in the most toxic GHS

371 acute categories, especially for the 3T3 NRU test method. For the 3T3 NRU test method, the

372 mean interlaboratory CV for chemicals in the classes for $LD_{50} \leq 5$ mg/kg (69%) and $5 < LD_{50}$

373    $\leq$ 50 mg/kg (78%) were much larger than the mean overall interlaboratory CV (46%,). For

374   the NHK NRU test method, the mean interlaboratory CV for chemicals in the classes for 5 <

375    $LD_{50} \leq$ 5 mg/kg (37%) and 5 < $LD_{50} \leq$ 50 mg/kg (41%) were much larger than the mean

376   overall interlaboratory CV (28%).

377

378   For the characteristics amenable to correlation analysis, none of the correlation coefficients

379   were large (absolute value of r_s < 0.5), but IC_{50} (p = 0.015) and boiling point (p = 0.021)

380   exhibited statistically significant correlations (p < 0.05) to interlaboratory CV for the 3T3

381   NRU test method. There was a negative correlation between interlaboratory CV and IC_{50},

382   but the correlation between boiling point and interlaboratory CV was positive. The positive

383   correlation of CV with boiling point was largely consistent with the categorical analysis of

384   volatility. The substances that exhibited volatile characteristics in the 3T3 NRU test method

385   had slightly higher mean CV than for the substances that did not exhibit volatile

386   characteristics (51 vs. 46%). For the NHK NRU test method, only IC_{50} was significantly

387   correlated (p = 0.024) to interlaboratory CV with a negative correlation (r_s = -0.271).

388

389   7.2.3   Comparison of Laboratory-Specific Linear Regression Analyses for the Prediction

390   of *In Vivo* Rodent LD_{50} Values from *In Vitro* NRU IC_{50} Values

391   The laboratory-specific regressions presented in **Table 6-1** of **Section 6.1.1** were compared

392   to one another (for each test method) with a goodness of fit F-test as described in **Section**

393   **5.3.3.** The comparisons indicated that the laboratory-specific regressions for both test

394   methods were not significantly different (p < 0.05) from one another. The comparison of the

395   laboratory-specific 3T3 NRU regressions to one another yielded p = 0.796. The comparison

396   of the laboratory-specific NHK NRU regressions to one another yielded p = 0.985. Because

397   the laboratory-specific regressions were not statistically different, data were combined into a

398   single regression for each test method using a geometric mean of the laboratory-specific IC_{50}

399   values for each substance (see **Section 6.1.1**).

400

401   7.2.4   Laboratory Concordance for the Prediction of GHS Acute Oral Toxicity Category

402   This section provides the percentage of substances for which the laboratory-specific IC_{50} data

403   yielded the same (for all three laboratories) GHS toxicity categorization when used with the

404   regressions evaluated in **Sections 6.3.1** through **6.3.3**. Data for the same reference

405   substances for each test method were evaluated to determine the laboratory concordance for

406   each regression. Forty-three substances were evaluated for the 3T3 NRU test method and 44

407   substances were evaluated for the NHK NRU test method. Of the original 72 substances

408   tested, epinephrine bitartrate, colchicine, and propylparaben were excluded from all analyses

409   because they were removed from the calculation of the RC rat-only weight regressions due to

410   the lack of rat oral reference LD_{50} data. The 21 substances with specific mechanisms of

411   toxicity in **Table 6-3** were excluded from all analyses to be consistent with those removed

412   from the RC rat-only weight regression excluding substances with specific mechanisms of

413   toxicity. These substances have known mechanisms of toxicity that are not expected to be

414   active in the 3T3 or NHK cell cultures. Carbon tetrachloride, methanol, gibberellic acid,

415   lithium carbonate, and xylene were excluded from the 3T3 NRU evaluations because at least

416   one laboratory failed to attain sufficient toxicity in any test for the calculation of an IC_{50}.

417   Carbon tetrachloride, methanol, 1,1,1-trichloroethane, and xylene were excluded from the

418   NHK NRU analyses because at least one laboratory failed to attain sufficient toxicity in any

419   test for the calculation of an IC_{50}.

420

421   *Laboratory Concordance for the 3T3 and NHK NRU Test Methods with the RC Millimole*

422   *Regression*

423   **Appendix J** (**Table J-1** for the 3T3 NRU test method and **Table J-3** for the NHK NRU test

424   method) shows the laboratory concordance of the observed (i.e., *in vivo* categories for the

425   initial LD_{50} values in **Table 3-2**) and predicted GHS toxicity categories for each substance

426   determined in each *in vitro* NRU cytotoxicity test method using the laboratory-specific

427   geometric mean IC_{50} values and the RC millimole regression, log LD_{50} (mmol/kg) = 0.435 x

428   log IC_{50} (mM) + 0.625. The observed LD_{50} values are the rodent LD_{50} values from **Table 3-**

429   **2.**

430

431   For the 43 substances that yielded IC_{50} results in all laboratories using the 3T3 NRU test

432   method, the laboratories agreed on the GHS toxicity category for 31 substances (72%). The

433   12 substances that produced discordant results among the laboratories were cupric sulfate

434   pentahydrate, cycloheximide, dimethylformamide, diquat dibromide, phenol, phenylthiourea,

435   sodium arsenite, sodium oxalate, sodium selenate, thallium sulfate, triethylenemelamine, and

436   1,1,1-trichloroethane. The laboratory predictions for these substances disagreed by one GHS

437   toxicity category.

438

439   For the 44 substances that yielded IC_{50} results in all laboratories using the NHK NRU test

440   method, the laboratories agreed on toxicity category for 39 substances (89%). The five

441   substances that produced discordant results among the laboratories were arsenic trioxide,

442   digoxin, ethanol, 2-propanol, and sodium arsenite. The laboratory predictions for these

443   substances disagreed by one toxicity category. Laboratory concordance was greater for the

444   NHK assay than for the 3T3 assay (89% vs 72%).

445

446   *Laboratory Concordance of the 3T3 and NHK NRU Test Methods with the RC Rat-Only*

447   *Weight Regression*

448   **Appendix J** (**Table J-5** for the 3T3 NRU test method and **Table J-6** for the NHK NRU test

449   method) shows the laboratory concordance of the observed (i.e., *in vivo* reference categories

450   for LD_{50} values in **Table 4-2**) and predicted GHS toxicity categories for each substance as

451   determined for each test method using the laboratory-specific geometric mean IC_{50} in the RC

452   rat-only weight regression, log LD_{50} (mg/kg) = log IC_{50} (μg/mL) x 0.372 + 2.024, from

453   **Table 6-2.**

454

455   For the 43 substances that yielded IC_{50} results in all laboratories using the 3T3 NRU test

456   method, the laboratories agreed on the GHS toxicity category for 34 substances (79%). The

457   nine substances that produced discordant results among the laboratories were boric acid,

458   cupric sulfate pentahydrate, cycloheximide, 2-propanol, propranolol HCl, sodium selenate,

459   thallium sulfate, triethylenemelamine, and 1,1,1-trichloroethane. The laboratory predictions

460   for these substances disagreed by one GHS toxicity category.

461

462   For the 44 substances that yielded IC_{50} results in all laboratories using the NHK NRU test

463   method, the laboratories agreed on toxicity category for 39 substances (89%). The five

464   substances that produced discordant results among the laboratories were arsenic trioxide,

465   digoxin, glycerol, sodium chloride, and thallium sulfate. The laboratory predictions for these

466   substances disagreed by one toxicity category. Laboratory concordance was greater for the

467   NHK assay than for the 3T3 assay (89% vs 79%).

468

469   *Laboratory Concordance of the 3T3 and NHK NRU Test Methods with the RC Rat-Only*

470   *Weight Regression Excluding Substances with Specific Mechanisms of Toxicity*

471   **Appendix J** (**Table J-7** for the 3T3 NRU test method and **Table J-8** for the NHK NRU test

472   method) shows the laboratory concordance of the observed (i.e., *in vivo*) and predicted GHS

473   toxicity categories for each substance as determined for each test method using the

474   laboratory-specific geometric mean IC_{50} values in the RC rat-only weight regression after

475   exclusion of substances with specific mechanisms of toxicity, log LD_{50} (mg/kg) = log IC_{50}

476   (μg/mL) x 0.357 + 2.194 (**Table 6-2**).

477

478   For the 43 substances considered in the analysis of the 3T3 NRU test method, the three

479   laboratories agreed on the toxicity category for 36 (84%) of the substances. The seven

480   substances that produced discordant results among the laboratories were boric acid, cupric

481   sulfate pentahydrate, diquat dibromide, sodium hypochlorite, thallium sulfate, 1,1,1-

482   trichloroethane, and valproic acid. The laboratory predictions for these substances disagreed

483   by one GHS toxicity category.

484

485   The extent of laboratory concordance for the RC rat-only weight regression after excluding

486   substances with specific mechanisms of toxicity was the same for the NHK NRU test method

487   (i.e., 84%, 37/44). The seven substances that produced discordant results among the

488   laboratories were arsenic trioxide, digoxin, glycerol, hexachlorophene, mercury chloride,

489   sodium chloride, and sodium hypochlorite. The laboratory predictions for these substances

490   disagreed by one GHS toxicity category.

491

492   **7.3   Historical Positive Control Data**

493

494   The reproducibility of the positive control (SLS) data was assessed by CV analysis,

495   ANOVA, and linear regression over time as described in **Section 5.3.4**. The SLS data

496   analyzed for variability are slightly different from those used to determine the PC acceptance

497   limits shown in **Table 5-2**. To get an assessment of the true variation of SLS IC_{50} values, the

498   reproducibility analyses included IC_{50} values from SLS tests that failed the test acceptance

499   criterion for the IC_{50} acceptance limits determined for each study phase. These additional

500   SLS tests, however, passed all other test acceptance criteria. If more than one SLS test was

501   performed in a single day (for each test method and laboratory), the IC_{50} values were

502   averaged to determine a single IC_{50} for the day so that multiple results from a single day

503   would not overly influence the average for each phase.

504

505   **Figure 7-1** shows the average SLS IC_{50} values for each test method, laboratory, and study

506   phase. Graphically, it appears that the SLS IC_{50} for the 3T3 NRU test method was relatively

507   consistent over the entire period of the study (approximately 2.5 years). The intralaboratory

508   CV values (shown in **Figure 7-1**) for the individual study phases ranged from 5% to 24%.

509   With the exception of the Phase Ib CV at FAL, the CV values for each laboratory and phase

510   were less than 20%. The interlaboratory CV values were even smaller: 6% for Phases Ia and

511   Ib; 10% for Phase II; and 2% for Phase III.

512

513   **Figure 7-1** shows that the SLS IC_{50} for the NHK NRU test method tended to vary with time,

514   but, with the exception of the SLS IC_{50} results from FAL, there appeared to be no consistent

515   trend. The IC_{50} values from FAL, which changed NHK cell culture methods after Phase Ib

516   (see **Section 5.1.3**), tended to decrease over time. Although the change in cell culture

517   methods reduced the magnitude of the IC_{50}, the variability (as evidenced by the

518   intralaboratory CV values shown in **Figure 7-1**) remained relatively high (CV $\geq$ 34% for all

519   FAL study phases). The CV values for all the laboratories and study phases indicated that

520   the SLS IC_{50} values for the NHK NRU test method was more variable within laboratories

521   than the SLS IC_{50} for 3T3 NRU test method. CV values for the SLS IC_{50} for the NHK NRU

522   test method ranged from 11 to 51%, with nine of the 12 values greater than 20%. The

523   interlaboratory CV values, which were also greater than those for the 3T3 NRU test method,

524   were: 39% for Phase Ia; 21% for Phase Ib; and 31% for Phase II; and 8% for Phase III.

525

526

526   **Figure 7-1**   **SLS IC_{50} for Each Laboratory and Study Phase**

527   a   3T3 NRU Test Method
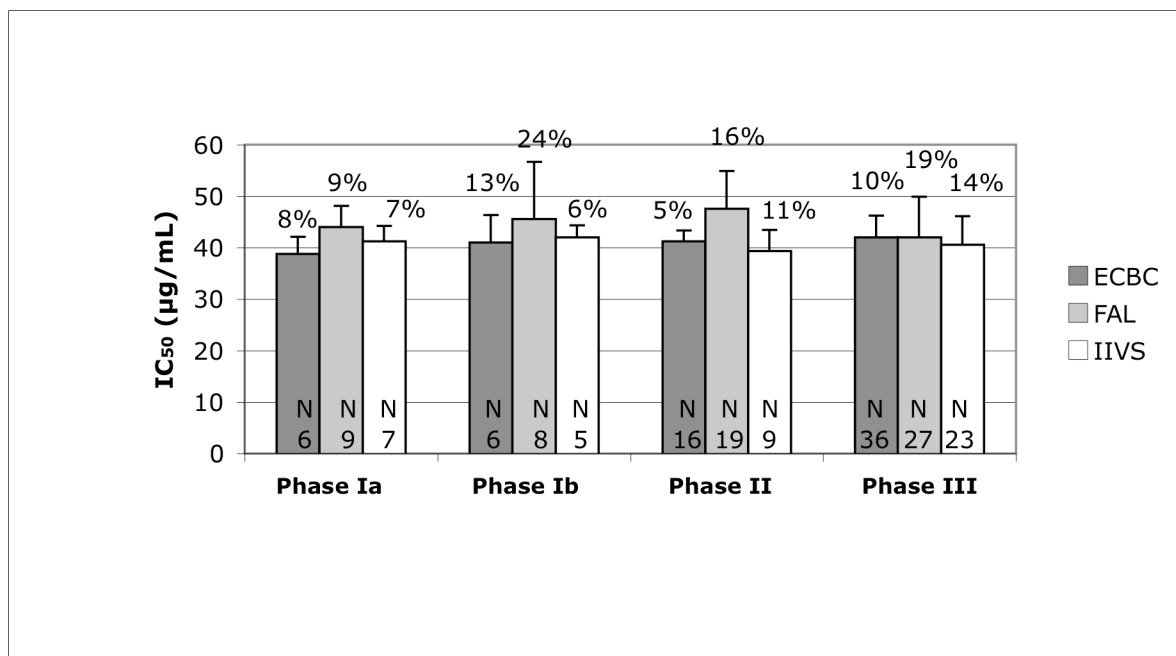
528

529   b   NHK NRU Test Method
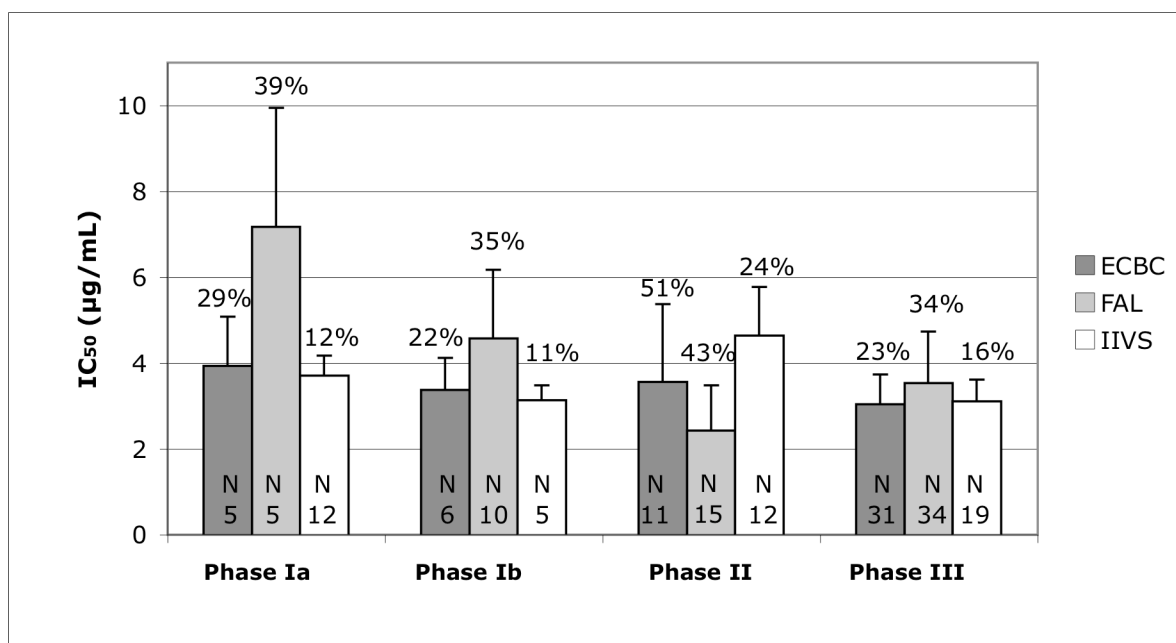
530

531   Bars show mean IC_{50} values. Error bars show standard deviation. Percent values above error bars are
532   intralaboratory CVs.
533   Laboratories: ECBC- U.S. Army Edgewood Chemical Biological Center; FAL – FRAME
534   Alternatives Laboratory; IIVS – Institute for In Vitro Sciences.
535

535   7.3.1   ANOVA and Linear Regression Results for the 3T3 NRU Test Method

536   *SLS IC_{50} Variation with Time*

537   **Table 7-9** shows the ANOVA results for SLS from the 3T3 NRU test method. When the

538   IC_{50} values within each laboratory were compared by study phase (i.e., the ANOVA factor

539   was study phase), there were no statistically significant differences ($p < 0.01$) between study

540   phases for any laboratory. **Table 7-10** shows that the slopes of the linear regressions of the

541   IC_{50} values over time (expressed as index values) were statistically different from zero for

542   ECBC and FAL ($p = 0.001$ and 0.012, respectively). Since the slopes were so small

543   (0.000204 and -0.000324), they were considered to be unimportant. The slope of the IIVS

544   regression of SLS IC_{50} over time was not statistically different from zero ($p = 0.651$; **Table**

545   **7-10**), which was entirely consistent with the ANOVA (**Table 7-9**) indicating that SLS IC_{50}

546   from IIVS did not vary with study phase ($p = 0.854$). The ANOVA with study phase as the

547   factor (with laboratories combined) indicated that the 3T3 NRU IC_{50} values from all the

548   laboratories were consistent over time since data from the various study phases were not

549   statistically significantly different ($p = 0.304$).

550

551   *Comparison of SLS IC_{50} Among the Laboratories*

552   When all study phases from each laboratory were combined, ANOVA, with laboratory as the

553   factor, indicated that the SLS IC_{50} for the 3T3 NRU test method differed in some statistically

554   significant fashion among the laboratories ($p < 0.006$). However, the differences between

555   laboratories look rather small in **Figure 7-1** since the SDs for the laboratories clearly overlap

556   one another.

557

558

559

560   **Table 7-9   ANOVA Results for SLS IC$_{50}$ from the 3T3 NRU Test Method**

Study Phase/ Laboratory	ECBC				FAL				IIVS			
	Log Mean IC$_{50}$ (mM)	SD	N	P[1]	Log Mean IC$_{50}$ (mM)	SD	N	P[1]	Log Mean IC$_{50}$ (mM)	SD	N	P[1]
Test for differences between phases within each laboratory												
Phase Ia	-0.876	0.042	6	0.031	-0.811	0.046	9	0.015	-0.850	0.034	7	0.854
Phase Ib	-0.864	0.066	6		-0.846	0.065	8		-0.838	0.025	5	
Phase II	-0.848	0.027	16		-0.796	0.057	19		-0.854	0.025	8	
Phase III	-0.842	0.036	36		-0.851	0.066	27		-0.844	0.041	23	
Test for differences between laboratories (phases combined)												
All Phases	-0.849	0.039	64	0.006	-0.826	0.062	63		-0.847	0.035	44	
Test for differences between phases (laboratories combined)												
Phase Ia	-0.839	0.049	22	0.304								
Phase Ib	-0.850	0.056	19									
Phase II	-0.831	0.047	34									
Phase III	0.845	0.045	86									

561   [1]Statistically significant at $p < 0.01$.
562   Abbreviations: N- number of values; SD – standard deviation. Laboratories: ECBC- U.S. Army Edgewood
563   Chemical Biological Center; FAL – FRAME Alternatives Laboratory; IIVS – Institute for In Vitro Sciences.
564

565   **Table 7-10   Linear Regression Analysis of SLS IC$_{50}$ Over Time[1]**

Test Method/ Laboratory	Slope	P-value (Slope)[2]	Intercept
3T3 NRU			
ECBC	0.000204	0.001	-0.874
FAL	-0.000324	0.012	-0.796
IIVS	0.0000304	0.651	-0.850
NHK NRU			
ECBC	-0.000559	0.002	-1.901
FAL	-0.00112	< 0.001	-1.737
IIVS	-0.000445	0.002	-1.885

566   [1]Time was expressed as index values. The index value of each test reflected the
567   order of testing without respect to the time lapsing between tests.
568   [2]Statistically significant from zero at p < 0.05.
569   Laboratories: ECBC- U.S. Army Edgewood Chemical Biological Center; FAL –
570   FRAME Alternatives Laboratory; IIVS – Institute for In Vitro Sciences.

571

572

573   7.3.2   ANOVA and Linear Regression Results for the NHK NRU Test Method

574   *SLS IC$_{50}$ Variation with Time*

575   **Table 7-11** shows the ANOVA results for the NHK NRU test method. When the IC$_{50}$ values

576   within each laboratory were compared by study phase (i.e., the ANOVA factor was phase),

577   the phases were statistically different (p < 0.01) at each laboratory. The IC$_{50}$ values from the

578   various study phases were also significantly different from one another when the laboratory

579   data were combined (p < 0.001). Linear regression analyses showed that the slopes for IC$_{50}$

580   over time (expressed as an index values) were statistically significantly greater than zero for

581   each laboratory (see **Table 7-10**). Since the slopes were so small (-0.000559, -0.00112, and -

582   0.000445), they were considered to be unimportant.

583   **Table 7-11   ANOVA Results for SLS IC$_{50}$ from the NHK NRU Test Method**

Study Phase/ Laboratory	ECBC				FAL				IIVS			
	Log Mean IC$_{50}$ (mM)	SD	N	P[1]	Log Mean IC$_{50}$ (mM)	SD	N	P[1]	Log Mean IC$_{50}$ (mM)	SD	N	P[1]
Test for differences between phases within each laboratory												
Phase Ia	-1.867	0.135	5	0.001	-1.656	0.125	5	< 0.001	-1.904	0.060	12	< 0.001
Phase Ib	-1.936	0.092	6		-1.829	0.141	10		-1.965	0.046	5	
Phase II	-2.007	0.109	11		-1.982	0.173	15		-1.863	0.058	12	
Phase III	-1.990	0.098	31		-1.941	0.113	34		-1.972	0.070	19	
Test for differences between laboratories (phases combined)												
All Phases	-1.971	0.113	53	< 0.001	-1.879	0.175	64		-1.924	0.073	48	
Test for differences between phases (laboratories combined)												
Phase Ia	-1.833	0.143	22	< 0.001								
Phase Ib	-1.891	0.125	21									
Phase II	-1.964	0.139	38									
Phase III	-1.971	0.100	84									

584   [1]Statistically significant at p < 0.01.
585   Abbreviations: N- number of values; SD – standard deviation. Laboratories: ECBC – U.S. Army Edgewood
586   Chemical Biological Center; FAL – FRAME Alternatives Laboratory; IIVS – Institute for In Vitro Sciences.
587

588   *Comparison of SLS IC$_{50}$ Among the Laboratories*

589   The ANOVA results, with laboratory as a factor (**Table 7-11)** indicated that the SLS IC$_{50}$

590   was statistically different among the laboratories when the data from the study phases were

591   pooled (p < 0.001). **Figure 7-1** shows that the SLS data from ECBC and IIVS were rather

592   similar for Phases Ia, Ib, and III. The SLS IC$_{50}$ data from FAL looks different from the other

593   two laboratories for Phases Ia, Ib, and II, but the bars and SDs for Phase III show that the

594   data from all laboratories were similar.

595

596   **7.4   Laboratory Concordance for Solvent Selection**

597

598   The solvents used to dissolve the reference substances are shown in **Table 7-12**. For Phases

599   Ib and II, the SMT selected the solvents to use for cytotoxicity testing based on the solubility

600   results provided by BioReliance (see **Table 5-7**) using the solubility protocol in **Appendix**

601   **G2**. Despite the fact that the solubility of an individual substance in 3T3 medium and NHK

602   medium might be different, the SMT chose the same solvent for both test methods, rather

603   than choosing one for the 3T3 assay and one for the NHK assay. For example, if solubility in

604   the 3T3 medium was ≥ 2 mg/mL and solubility in the NHK medium was < 2 mg/mL, and the

605   substance was soluble in DMSO at 200 mg/mL, then the SMT selected DMSO as the solvent

606   for cytotoxicity testing.

607

608   During Phases Ib and II, the SMT noted that BioReliance sometimes achieved higher

609   solubility than the cytotoxicity laboratories (e.g., see the results for arsenic trioxide,

610   aminopterin, and chloramphenicol in **Table 5-7**). In an attempt to avoid the selection of a

611   solvent for which one or more laboratories could not achieve the desired solubility, the SMT

612   used the solubility data from all the laboratories to determine solvent selections for

613   cytotoxicity testing in Phase III. The SMT viewed BioReliance's NHK and 3T3 media

614   solubility results for each substance in Phases Ib and II to be one result for media and took a

615   similar approach in Phase III when considering all the laboratory results to determine the

616   solvent to use for cytotoxicity testing. For example, if one laboratory had achieved solubility

617   at 2 mg/mL in medium, but the other laboratories had not, and the substance was soluble in

618   DMSO at 200 mg/mL, then the SMT selected DMSO as the solvent. **Table 7-12** shows that

619   cell culture medium was used to test as the solvent for 38 substances and DMSO was used as

620   the solvent for 34 substances.

621

622   The solubility protocol used by the cytotoxicity laboratories failed to guide the selection of a

623   solvent for five substances because they were insoluble at all concentrations tested in at least

624   one laboratory. Arsenic trioxide was insoluble at all the cytotoxicity laboratories. IIVS also

625   found sodium oxalate, strychnine, and triethylenemelamine insoluble in any solvent, and

626   FAL found thallium sulfate insoluble in any solvent. To select a solvent for cytotoxicity

627   testing of these substances, the SMT used the solubility results from the laboratories that did

628   achieve solubility.

629

Table 7-12   Solvent Determinations by Laboratory

Reference Substance	Solvent for Testing[1]	ECBC	FAL	IIVS
Acetaminophen	DMSO	Medium	Medium	DMSO
Acetonitrile	Medium	Medium	Medium	Medium
Acetylsalicylic acid	DMSO	Medium	DMSO	Medium
Aminopterin	DMSO	DMSO	DMSO	DMSO
5-Aminosalicylic acid	Medium	Medium	Medium	Medium
Amitriptyline HCl	DMSO	DMSO	DMSO	DMSO
Arsenic III trioxide	Medium	ID	ID	ID
Atropine sulfate	Medium	Medium	Medium	Medium
Boric aid	Medium	Medium	Medium	Medium
Busulfan	DMSO	DMSO	DMSO	DMSO
Cadmium II chloride	DMSO	DMSO	DMSO	DMSO
Caffeine	Medium	Medium	Medium	Medium
Carbamazepine	DMSO	Medium	DMSO	DMSO
Carbon tetrachloride	DMSO	Medium	DMSO	Medium
Chloral hydrate	Medium	Medium	Medium	Medium
Chloramphenicol	DMSO	DMSO	DMSO	Medium
Citric acid	Medium	Medium	Medium	Medium
Colchicine	Medium	Medium	Medium	Medium
Cupric sulfate pentahydrate	Medium	Medium	Medium	Medium
Cycloheximide	Medium	Medium	Medium	Medium
Dibutyl phthalate	DMSO	DMSO	DMSO	DMSO
Dichlorvos (DDVP)	DMSO	Medium	DMSO	Medium
Diethyl phthalate	DMSO	DMSO	DMSO	DMSO
Digoxin	DMSO	DMSO	DMSO	DMSO
Dimethylformamide	Medium	Medium	Medium	Medium
Diquat dibromide monohydrate	Medium	Medium	Medium	Medium
Disulfoton	DMSO	DMSO	DMSO	DMSO
Endosulfan	DMSO	DMSO	DMSO	DMSO
Epinephrine bitartrate	Medium	Medium	Medium	Medium
Ethanol	Medium	Medium	Medium	Medium
Ethylene glycol	Medium	Medium	Medium	Medium
Fenpropathrin	DMSO	DMSO	DMSO	DMSO

Table 7-12   Solvent Determinations by Laboratory

Reference Substance	Solvent for Testing[1]	ECBC	FAL	IIVS
Gibberellic acid	Medium	Medium	Medium	Medium
Glutethimide	DMSO	DMSO	DMSO	DMSO
Glycerol	Medium	Medium	Medium	Medium
Haloperidol	DMSO	DMSO	DMSO	DMSO
Hexachlorophene	DMSO	DMSO	DMSO	DMSO
Lactic acid	Medium	Medium	Medium	Medium
Lindane	DMSO	DMSO	DMSO	DMSO
Lithium I carbonate	Medium	Medium	Medium	Medium
Meprobamate	DMSO	Medium	Medium	DMSO
Mercury II chloride	DMSO	DMSO	DMSO	DMSO
Methanol	DMSO	Medium	Medium	DMSO
Nicotine	Medium	Medium	Medium	Medium
Paraquat	Medium	Medium	Medium	Medium
Parathion	DMSO	DMSO	DMSO	DMSO
Phenobarbital	DMSO	Medium	DMSO	DMSO
Phenol	Medium	Medium	Medium	Medium
Phenylthiourea	DMSO	DMSO	Medium	DMSO
Physostigmine	DMSO	Medium	DMSO	DMSO
Potassium I chloride	Medium	Medium	Medium	Medium
Potassium cyanide	Medium	Medium	Medium	Medium
Procainamide HCl	Medium	Medium	Medium	Medium
2-Propanol	Medium	Medium	Medium	Medium
Propranolol HCl	DMSO	Medium	Medium	Medium
Propylparaben	DMSO	DMSO	DMSO	DMSO
Sodium arsenite	Medium	Medium	Medium	Medium
Sodium chloride	Medium	Medium	Medium	Medium
Sodium dichromate dihydrate	Medium	Medium	Medium	Medium
Sodium fluoride	Medium	Medium	Medium	Medium
Sodium hypochlorite	Medium	Medium	Medium	Medium
Sodium oxalate	Medium	Medium	Medium	ID
Sodium selenate	Medium	Medium	Medium	Medium
Strychnine	Medium	Medium	Medium	ID
Thallium I sulfate	Medium	Medium	ID	Medium
Trichloroacetic acid	Medium	Medium	Medium	Medium
1,1,1-Trichloroethane	Medium	Medium	Medium	Medium
Triethylenemelamine	DMSO	Medium	DMSO	ID
Triphenyltin hydroxide	DMSO	DMSO	DMSO	DMSO
Valproic acid	DMSO	Medium	DMSO	DMSO
Verapamil HCl	DMSO	DMSO	DMSO	DMSO
Xylene	DMSO	DMSO	DMSO	DMSO
DMSO Total	34	22	29	28
Medium Total	38	49	41	40

630   ID-insufficient data to select solvent.
631   [1]Solvents for testing as determined by the SMT and used in the study by each laboratory: Medium = cell culture
632   medium; DMSO = dimethyl sulfoxide
633   ECBC – US Army Edgewood Chemical Biological Center; FAL – FRAME Alternatives Laboratory; IIVS –
634   Institute for In Vitro Sciences
635
636

636   The cytotoxicity laboratories selected the same solvent for 55 of the 72 reference substances

637   (76%). Excluding the five substances that were found to be insoluble in any solvent by at

638   least one laboratory, there were 12 substances for which the cytotoxicity laboratories

639   disagreed: acetaminophen, acetylsalicylic acid, carbamazepine, carbon tetrachloride,

640   chloramphenicol, dichlorvos, meprobamate, methanol, phenobarbital, phenylthiourea,

641   physostigmine, and valproic acid. Every laboratory reported relatively low solubility, ≤ 2

642   mg/mL, in medium for these substances. Since 2 mg/mL in medium is the departure point

643   for the selection medium or DMSO, a small variation in results causes the laboratories to

644   select different solvents. The solubility of acetaminophen, for example was reported as 2

645   mg/mL in culture media by ECBC and FAL, but < 2 mg/mL by IIVS. IIVS found it soluble

646   in 200 mg/mL DMSO and selected DMSO as the solvent. ECBC and FAL selected the

647   culture media as the solvent. The SMT selected DMSO as the solvent for acetaminophen to

648   be used by all laboratories.

649
650   **7.5   Summary**

651

652   Intra- and inter-laboratory reproducibility were assessed using ANOVA, CV analysis,

653   comparison of the laboratory-specific IC_{50}-LD_{50} regressions to one another (for each test

654   method) and laboratory concordance for the GHS acute oral toxicity category predictions.

655   ANOVA permits statistical comparisons of laboratories and experimental averages, while

656   controlling for other factors. CV analysis is an empirical way of expressing the relative

657   magnitudes of variability on a standardized scale. ANOVA results for the reference

658   substances showed significant laboratory differences for 26 substances for the 3T3 NRU test

659   method and seven substances for the NHK test method. Intralaboratory CV values were 1-

660   122% for the 3T3 NRU test method and 1-129% for the NHK NRU test method. Mean

661   intralaboratory CV values were 26% for both test methods, but the NHK NRU test method

662   had a lower interlaboratory CV (28% vs 46%). Interlaboratory CV values were 2-135% for

663   the 3T3 NRU test method and 1-99% for the NHK NRU test method. FAL had the highest

664   mean intralaboratory CV for both test methods (33% for the 3T3 NRU test method and 42%

665   for the NHK NRU test method).

666

667   An analysis to determine the relationship between the chemical attributes and interlaboratory

668   CV indicated that physical form, solubility, and volatility had little effect on CV. CV seemed

669   to be related, however, to chemical class, GHS acute toxicity category, IC_{50}, and boiling

670   point. Reference substances in the amide class had unusually low mean interlaboratory CV

671   values for both the 3T3 NRU test method (15%) and NHK NRU test method (16%)

672   compared with the overall mean interlaboratory CV values (46% for the 3T3 NRU test

673   method and 28% for the NHK NRU test method). Reference substances in the

674   organophosphate and heterocyclic classes had unusually high mean interlaboratory CV

675   values for the 3T3 NRU test method (74% and 71%, respectively), but not for the NHK NRU

676   test method. Mean interlaboratory CV values were large for substances in the most toxic

677   GHS acute categories, especially for the 3T3 NRU test method. The mean interlaboratory

678   CV for substances in the classes for $LD_{50} \leq 5$ mg/kg (69%) and $5 < LD_{50} \leq 50$ mg/kg (78%)

679   were larger than the mean overall interlaboratory CV (46%,) for the 3T3 NRU test method.

680   For the NHK NRU test method, the mean interlaboratory CV was 37% for substances with

681    $LD_{50} \leq 5$ mg/kg and 41% for substances with $5 < LD_{50} \leq 50$ mg/kg while the mean overall

682   interlaboratory CV was 28%. A Spearman correlation analysis indicated that IC_{50} was

683   negatively correlated to interlaboratory CV for both 3T3 ($p = 0.015$) and NHK ($p = 0.024$)

684   NRU test methods and that boiling point was positively correlated to interlaboratory CV ($p =$

685   0.021) for the 3T3 NRU test method.

686

687   The analysis of interlaboratory reproducibility by evaluating the similarity of the laboratory

688   specific IC_{50}-LD_{50} regressions indicated that the laboratory regressions for both test methods

689   were not significantly different ($p < 0.05$) from one another ($p = 0.796$ for the 3T3 NRU and

690    $p = 0.985$ for the NHK NRU). The evaluation of laboratory concordance for the prediction

691   of GHS acute oral toxicity category when the laboratory-specific IC_{50} data were applied to

692   the same regression yielded the following proportions of substances for which all laboratories

693   agreed on the GHS acute oral toxicity categorization:

694   • 78% (52/67) for the 3T3 NRU and 87% (59/68) for the NHK NRU with the RC

695   regression

696   • 81% (52/64) for the 3T3 NRU and 91% (59/65) for the NHK NRU with the RC

697   rat only weight regression

698   • 84% for the both test methods (36/43 for the 3T3 NRU and 37/44 for the NHK

699   NRU) with the RC rat only weight regression excluding substances with

700   specific mechanisms of action

701

702   ANOVA results for the positive control, SLS, IC_{50} in the 3T3 NRU test method indicated

703   that there were significant differences among laboratories ($p = 0.006$) and but not between

704   study phases within laboratories ($p > 0.01$). However, interlaboratory CV values, which

705   ranged from 2% to 10% for the study phases, indicated that the laboratories were similar.

706   Intralaboratory CV values for the study phases ranged from 5% to 24%. SLS IC_{50} values for

707   the NHK NRU test method were more variable than those for the 3T3 NRU test method.

708   ANOVA results for SLS in the NHK NRU test method indicated that there were significant

709   differences between laboratories ($p < 0.001$) and between study phases within laboratories (p

710   ≤ 0.001). A change in cell culture methods at FAL decreased the SLS IC_{50} from Phase Ib to

711   Phase II. Intralaboratory CV values for the NHK NRU SLS IC_{50} during the various study

712   phases ranged from 11% to 51%. Interlaboratory CV values for SLS in the NHK NRU test

713   method ranged from 8% to 39%.

714

715   Cell culture medium was used as the solvent for testing 38 substances and DMSO was used

716   for 34 substances. The laboratory concordance in selecting solvent for the reference

717   substances using the solubility protocol was 76% (55/72).

718

718
719
720
721
722
723
724
725
726
727
728
729
730
731
732   ***[This Page Intentionally Left Blank]***
733
734

21
22
23
24
25
26
27
28
29
30
31
32
33
34
35   ***[This Page Intentionally Left Blank]***
36

36   **8.0   3T3 AND NHK NRU TEST METHOD DATA QUALITY**

37

38   This section of the BRD presents the extent of adherence to national and international GLP

39   guidelines during for generation of the NICEATM/ECVAM validation study data. Data

40   quality is described along with any deviations from the guidelines and the impact of any

41   noncompliance. Statistical results are provided to show comparison of data generation,

42   collection, and reporting of the two GLP adherent cytotoxicity testing laboratories and the

43   one non-GLP adherent cytotoxicity testing laboratory as well as the GLP laboratory that

44   distributed the reference substances and performed solubility studies. Discussions of various

45   quality assurance aspects of the study are included.

46

47   **8.1   Adherence to Good Laboratory Practice Guidelines**

48

49   8.1.1   <u>Guidelines Followed for *In Vitro* NRU Cytotoxicity Testing</u>

50   *Good Laboratory Practices*

51   The SOW provided the following definition of U.S. Regulatory agency GLPs to each

52   laboratory:

53   *"Regulations governing the conduct, procedures, and operations of toxicology*

54   *laboratories; regulations to assure the quality and integrity of the data and to address*

55   *such matters as organization and personnel, facilities, equipment, facility operations, test*

56   *and control articles, and validation study protocol, and conduct (U.S. Food and Drug*

57   *Administration, Title 21 CFR Part 58; U.S. Environmental Protection Agency, Title 40*

58   *CFR Part 160)."*

59

60   IIVS, ECBC, and BioReliance performed testing under all GLP guidelines. The details of

61   GLP compliance and training are addressed in **Section 11**.

62

63   *Spirit of GLP*

64   The SMT determined a definition for "spirit of GLP" and provided the following verbiage to

65   the laboratories:

66   *"Laboratories that are non GLP-compliant shall adhere to GLP principles and other*

67   *method parameters as put forth in this Statement of Work and the Test Method Protocols*

68   *(provided by NIEHS/NICEATM); documentation and accountability shall be equal to*

69   *GLP requirements; laboratories must make assurances that they are equal in*

70   *performance criteria and that there is parity amongst the laboratories."*

71

72   FAL performed testing in the "spirit of GLP" (see **Section 11.2.2**) by following the

73   international GLP standards referenced in the ECVAM Workshop 37 Report (Cooper-

74   Hannan 1999) and the OECD Principles of GLP (OECD 1998). The laboratory did not have

75   data and test method procedures reviewed by an independent quality assurance (QA) auditor.

76   At a minimum, the SOW directed FAL to routinely document the following laboratory tasks

77   (e.g., equipment monitoring) and record keeping (see **Table 8-1**) and to archive the

78   documents. The FAL laboratory already had most of the following procedures and

79   guidelines in place for routine laboratory procedures before initiation of this study. The

80   various general laboratory-related activities were documented in workbooks and logbooks

81   and the information was made available to the SMT.

82

83   **Table 8-1   SMT-Recommended Documentation for FAL Laboratory**

Daily	Per Use	Periodic
Temperatures Laboratory, incubators, water baths, refrigerators, freezers	Cryogenic Storage Unit Liquid N_2 volume	Laboratory Supplies[1] Lot numbers and expiration dates for stock media formulations and components, NRU reagents, tissue culture plasticware
Humidity/CO_2 Cell culture incubators	Equipment Calibration Balances, pH meters, and cell counters	Cells Quantity and cryogenic storage conditions for 3T3 and NHK cells
Visual Observations Cell Culture Growth	Reagents Lot numbers and expiration dates of medium/supplements	Equipment Calibration Incubators, laminar flow hoods, autoclaves, micropipettors, spectrophotometer plate readers, computers (software)

84   [1]Periodic documentation for laboratory supplies occurs when supplies are purchased and received in the
85   laboratory
86

87

88   *Good Cell Culture Practices (GCCP)*

89   The SMT provided guidance in the SOW for implementing GLPs in a cell culture laboratory

90   environment. The initial assumption by the SMT was that each laboratory had the basic cell

91   culture skills and knowledge (e.g., as described in Freshney 2000) to perform the NRU

92   cytotoxicity test methods in a reliable manner. Reviews of historical documents and

93   scientific and professional exchanges with the laboratory personnel assured the SMT that

94   each laboratory had demonstrated, through previous validation studies and other scientific

95   endeavors, that personnel were capable of providing quality scientific data through the use of

96   good cell culture practices. A comparison of the SOW and the *in vitro* NRU cytotoxicity

97   protocols to the ECVAM Good Cell Culture Practices (GCCP) Reports (Hartung 2002;

98   Coecke et al. 2005) and the OECD document on GLPs and *in vitro* studies (OECD 2004a)

99   showed that the guidelines in place for the NICEATM/ECVAM study were harmonious with

100   the ECVAM and OECD guidelines.

101

102   8.1.2   Quality Assurance (QA) for *In Vitro* NRU Cytotoxicity Test Data

103   *Coded Reference Substances*

104   BioReliance acquired 73 high purity chemicals (72 reference substances and one positive

105   control chemical at 99% or greater purity when economically feasible) from reputable

106   commercial sources according to the SOW provided by the SMT (see **Appendix G**). Seven

107   reference substances were less than 99% pure (three less than 98% pure; lactic acid had the

108   lowest purity [89%]). The substances were coded with unique identification numbers and

109   provided to the testing laboratories in a blinded fashion. Preparation of substances for

110   distribution was performed under GLP guidelines. **Section 3.6** provides detailed information

111   concerning acquisition and distribution of reference substances.

112

113   *Solubility Testing and Data Review*

114   All laboratories performed solubility tests on all reference substances using the solvents and

115   procedures specified by the protocols provided by the SMT and submitted solubility data as

116   hard copy printouts and electronic worksheets. The laboratories also maintained solubility

117   data in their workbooks. The Study Directors reviewed all laboratory procedures and all data

118   produced at their respective laboratories. The QA designee reviewed all data in the GLP-

119   adherent laboratories. The SMT Project Coordinators served as informal QA reviewers for

120   the FAL (i.e., reviewed all raw data sheets). Detection of errors and omissions were reported

121   to FAL and corrections were requested. The SMT reviewed all solubility data and all NRU

122   assay data produced by all laboratories for this study.

123

124   The SMT reviews of submitted data in Phases Ia and Ib revealed that even after data review

125   by the Study Directors, data files contained an unacceptable high frequency of errors (see

126   **Section 2.6.3**). The laboratories were alerted to the problem and personnel from all the

127   laboratories attended a weeklong training session to enhance harmonization among the

128   laboratories. After the training, errors were still found in data files submitted for Phase III,

129   albeit less frequently; such errors generally occurred due to the rapid submission of data files

130   to the SMT shortly after the conclusion of each test. The formal QA review of the files

131   occurred later in each phase of the study.

132

133   Errors included typographical mistakes, transcriptional and data entry errors in the

134   Microsoft® EXCEL® and the GraphPad PRISM® 3.0 templates, and incorrect labeling of

135   files. The SMT reviewed every electronic file and hard copy printout throughout the study

136   and alerted the Study Directors when errors were found. All data files were checked for

137   consistency within the documents and for compliance with the protocols. The SMT also

138   documented errors on the hard copy printouts as handwritten notations and included these

139   notations in the electronic data summary files compiled for data management. Files that

140   were revised and/or corrected by the Study Director were resubmitted to the SMT and noted

141   as corrected files.

142

143   *In Vitro NRU Cytotoxicity Test Tallies*

144   Periodically, the laboratories received individualized test tallies from NICEATM that

145   detailed:

146   • the number of range finder tests performed

147   • the number of definitive tests performed and the pass/fail status of each test

148   • the number of positive control assays performed and the pass/fail status of each

149   test

150   •   the number of acceptable tests completed per the SMT and protocol

151   requirements

152   •   the status of test completion for each substance (i.e., whether one range finder

153   test and three acceptable definitive tests had been completed for the substance)

154

155   The laboratories compared the NICEATM tallies to their own records to verify consistency

156   and accuracy. Discrepancies were resolved through direct communication between the Study

157   Director and the SMT.

158

159   8.1.3   Guidelines Followed for *In Vivo* Rodent Oral LD_{50} Data Collection

160   The *in vitro* NRU cytotoxicity test methods are proposed as methods to predict starting doses

161   for acute oral lethality *in vivo* (specifically, rat) assays and not as replacement tests for an *in*

162   *vivo* reference method. No *in vivo* tests were performed for this validation study. All *in vivo*

163   data (i.e., rodent [rat and mouse] LD_{50} values) were collected by NICEATM through

164   reviews of the literature. All data and pertinent information were gathered and stored in a

165   spreadsheet database.

166

167   *Rodent Acute Oral LD_{50} Values Used in the Registry of Cytotoxicity (RC)*

168   The RC rodent (rat and mouse) acute oral LD_{50} values came largely from the 1983/84

169   RTECS® database (compiled by NIOSH). The RC is a database of acute oral LD_{50} values for

170   rats and mice obtained from RTECS® and IC_{50} values from *in vitro* cytotoxicity assays using

171   multiple cell lines and cytotoxicity endpoints for chemicals with known molecular weights

172   (Halle 1998). Collection and reporting methods used for generating the data were not a part

173   of any data collection hierarchy employed by the NIOSH. The data in the RTECS® database

174   were not evaluated for quality and accuracy by NIOSH. Many sources of the values come

175   from secondary references with no citation for the original report. GLP guidelines for acute

176   oral toxicity testing were not part of any criteria for determining acceptable data for the

177   database. The only criterion the NIOSH used for reporting acute oral toxicity data in

178   RTECS® was that the LD_{50} value was the most toxic LD_{50} value for a chemical that could be

179   found in the literature.

180

181   *Rodent Acute Oral LD$_{50}$ Values Collected by NICEATM*

182   One critical aspect of the study design was the establishment of a rat acute oral LD$_{50}$

183   reference value for each of the 72 reference substances (see **Section 4**). These reference

184   values were used to evaluate the extent to which the two *in vitro* test methods can predict rat

185   acute oral LD$_{50}$ values. Primary rat acute oral LD$_{50}$ studies were located through searching

186   electronic databases, published literature, and secondary references. Rat data were not

187   available for three of the reference substances and, for these, mouse acute oral LD$_{50}$ values

188   were collected. Very little data collected from the literature were produced under GLP

189   guidelines; in fact, only seven of the 455 LD$_{50}$ values collected were obtained under GLP

190   conditions.

191

192   **8.2   Results of Data Quality Audits**

193

194   The QA unit or designee of each GLP laboratory provided a systematic and critical

195   comparison of the data provided in the study report to the raw data in the laboratory records.

196   The SOW provided to each laboratory contained the following guidance on QA statements:

197   *"The Final Reports for all phases of the Validation Study shall be audited by the Quality*

198   *Assurance unit of the Testing Facility for GLP compliance and a QA Statement shall be*

199   *provided by the Testing Facility. Each Final Report shall identify: 1) the phases and*

200   *data inspected, 2) dates of inspection, and 3) dates findings were reported to the Study*

201   *Director and Testing Facility management. The QA Statement shall identify whether the*

202   *methods and results described in the Final Report accurately reflect the raw data*

203   *produced during the Validation Study."*

204

205   8.2.1   QA Statements

206   The QA statements from the GLP-compliant laboratories noted the QA reviews of:

207   • protocols

208   • laboratory standard operating procedures (SOPs)

209   • laboratory operations

210   • 3T3 and NHK NRU experiment data

211   • final report

212

213   The QA statements report that the test methods described in the protocols are the methods

214   that the laboratory personnel used and that the data reported to the SMT is an accurate

215   reflection of the raw data obtained by the laboratory. See **Section 8.2.2** for information about

216   the QA statements for the non-GLP laboratory.

217

218   8.2.2   QA Statements from the Laboratories

219   *BioReliance QA Statements*

220   The Study Director/Laboratory Director provided the following statement in all of the final

221   reports from BioReliance:

222   *"The solubility studies, acquisition, preparation, and distribution of the test chemicals*

223   *were conducted in compliance with GLP. Although not audited (per SOW), the work*

224   *described in this report for Phase X (i.e., Ia, Ib, and II) fully and accurately reflects to the*

225   *best of my knowledge the raw data generated in the study."*

226

227   *FAL QA Statements*

228   The Study Director for the FAL laboratory performed the final review of all data and reports

229   before sending to the SMT and provided two statements in the final reports (provided to the

230   SMT).

231   •   *"The laboratory worked under the principles of GLP whilst not being a GLP-*
232   *compliant laboratory."*

233   •   *"The report accurately reflects the work undertaken and the results obtained at*
234   *the FRAME Alternatives Laboratory."*

235

236   Since the SMT performed QA reviews of the FAL as an informal reviewer, formal QA

237   statements were not provided to FAL.

238

239   *ECBC QA Statements*

240   The QA statements reported what particular study phase and which laboratory procedures

241   were examined for compliance with GLP guidelines. In addition, the statement reiterated

242   that the scope of work, associated protocols, and quality control acceptance criteria were

243   updated/changed during the study which made it more difficult to assess the procedures and

244   data for conformance to the protocols. However, during the review of SOPs and the

245   observance of operations, the requirements and intent of GLP guidelines were continually

246   assessed. The QA reviews found the ECBC protocols to be in compliance with the

247   NICEATM/ECVAM study protocols. The phases of the studies inspected by the QA

248   designee were as follows:

249   • review of protocols and laboratory SOPs

250   • review of waste handling

251   • review of laboratory operations

252   • certification of new personnel

253   • review of data

254   • review of the final report for each phase

255

256   The QA designee also observed preparation of reference substances, 96-well plate

257   configuration, application of reference substance, annotation to the workbook, and

258   appropriate sterile technique while performing the testing. The number of inspections of

259   laboratory operations were reduced in the latter phases of the validation study since the same

260   personnel conducted the testing throughout the entire study.

261

262   ECBC Review Dates of Various Aspects of the Study

263   • Phase Ia: July 2002 through May 2003

264   • Phase Ib: July 2002 through January 2003

265   • Phase II: May 2003 through February 2004

266   • Phase III: November 2003 through March 2005

267

268   *IIVS QA Statements*

269   Because the IIVS QA unit is small, it carried out reviews in phases. The IIVS QA Statement

270   reads:

271   "This study has been divided into a series of in-process phases. Using a random

272   sampling approach, Quality Assurance monitors each of these phases over a series of

273   studies. Procedures, documentation, equipment records, etc., are examined to assure

274   that the study is performed in accordance with the U.S. FDA Good Laboratory

275   Practice regulations (21 CFR 58), the U.S. EPA GLP Standards (40 CFR 792 and 40

276   CFR 160) and the OECD Principles of Good Laboratory Practice and to assure that

277   the study is conducted according to the protocol and relevant Standard Operating

278   Procedures."

279

280   The phases of the studies inspected by the QA designee were as follows:

281   • protocol and initial paperwork

282   • reading of the plates (definitive assay)

283   • dilution of the test articles (definitive assay)

284   • termination of treatment and addition of the NR dye (definitive assay)

285   • cell concentration determination and seeding of the plates (third definitive)

286   • termination of treatment and addition of the NR dye

287   • washing the cells

288   • treatment of the cells

289   • draft report and data

290   • final report

291

292   <u>IIVS Review Dates of Various Aspects of the Study</u>

293   • Phase Ia: August 2002   Final Report Review: October 2005

294   • Phase Ib: January 2003   Final Report Review: October 2005

295   • Phase II: July-August 2003   Final Report Review: October 2005

296   • Phase III: January-November 2004 Final Report Review: October 2005

297

298   *Other QA Information*

299   Data generated by the laboratories and reviewed by their respective Study Directors were

300   provided directly to the SMT. Often, the data were provided electronically within days of the

301   end of testing. The SMT was very active as a secondary QA reviewer concerning all

302   information provided by the Study Directors. If the SMT found discrepancies, then the

303   Project Coordinators corresponded with the appropriate Study Director to rectify the mistake.

304   The Study Director made corrections/adjustments to any discrepancies in data reporting and

305   presented any changes to the SMT. The SMT did not initiate any external data quality audits.

306

307   The quality of the reference substances was assured in the form of certificates of analysis

308   provided by the chemical manufacturer to BioReliance at the time of purchase. The SMT

309   and the laboratories obtained certificates of analysis from CAMBREX specifically for

310   Clonetics® NHK culture medium and supplements. In addition, the SMT obtained quality

311   control data directly from CAMBREX technical departments for determining the NHK

312   medium's ability to support keratinocyte growth.

313

314   **8.3   Impact of Deviations from GLPs/Non-compliance**

315

316   Several error rates were determined by the SMT in regard to documentation, testing methods,

317   and data manipulation by the laboratories. Many errors (particularly in Phases Ia and Ib)

318   were minor mistakes (e.g., typographical, mislabeling) and did not affect the quality of the

319   data.

320

321   8.3.1   Laboratory Error Rates

322   During Phases Ia and Ib, the SMT was concerned about the number of errors in

323   documentation and testing methods and compiled the number of detected errors from each

324   laboratory. The types of errors noted and compiled included errors in documentation (e.g.,

325   reference substance identification did not match on all associated data sheets, IC_{20} and IC_{80}

326   values were switched in the EXCEL® template, a test acceptance criterion flag in data sheet

327   was incorrect, etc.) and in testing (e.g., wrong dilution scheme was used for the PC, wrong

328   SLS IC_{50} was used as the PC IC_{50}, etc.). Error rates were compiled as number of tests with

329   errors per total number of tests. As shown in **Table 2-3**, FAL had the highest error rates:

330   93% for the 3T3 assay and 41% for the NHK assay. The highest error rates of the other

331   laboratories were 10% for the 3T3 assay and 23% for the NHK assay (both ECBC).

332

333   There were very few errors detected in the Phase III data files. The SMT did not compile

334   typographical and transcriptional errors but reported the errors directly to the appropriate

335   Study Director so that the data sheets could be immediately rectified. The SMT did not

336   detect errors in the raw optical density data from the 96-well plates provided in each data file.

337   The laboratories and the SMT corrected any typographical and transcriptional errors (e.g.,

338   incorrect logIC$_{50}$ value entered) in the EXCEL® templates. The template formulas calculated

339   the correct values for the statistical analyses and the quality of the data was not

340   compromised.

341

342   For Phase III, assessment of error rates was performed specifically for Phase III for one

343   particular clerical error – the transfer of statistical results (e.g., ICx values) from the

344   GraphPad PRISM® 3.0 template to the Microsoft® EXCEL® template. It was often necessary

345   for the SMT to revise the Microsoft® EXCEL® data files provided by the laboratories

346   because the incorrect values had been transferred to the template. The SMT revised files

347   (using the data in the PRISM® 3.0 template) due to this error and reports as follows as the

348   number of errors/total number of definitive tests:

349

350   **Table 8-2   Error Rates**

Laboratory	Number of Errors Detected[1]	Number of Definitive Tests	Percentage of Tests with Detected Errors
ECBC	49	402	12
FAL	171	513	33
IIVS	25	419	6

351   [1]Clerical error – transfer of statistical results from PRISM® to EXCEL®

352

353   8.3.2   <u>Test Failure Rates for Definitive Tests and PC Tests</u>

354   **Table 8-3** illustrates the test failure rates experienced for Phase III of the validation study.

355   Approximately 25% of all 3T3 definitive tests and 18% of all NHK definitive tests failed

356   (i.e., did not meet test acceptance criteria). If a definitive test (see **Section 2.2.2** for the

357   definition of a definitive test) failed, then the laboratory repeated the test and attempted to

358   reach the goal of three acceptable definitive tests for each reference substance and each cell

359   type (see **Section 2.5** for criteria for repeating tests). PC failure occurred 0 – 18% of the time

360   with an overall average failure rate of 8% combined for both assays. FAL had the highest

361   individual laboratory test failure rates for 3T3 definitive tests (30%), NHK definitive tests

362   (32%), and NHK PC tests (18%). ECBC had the highest failure rate for 3T3 PC tests (11%).

363

364   Phase III guidelines called for each laboratory to provide three acceptable definitive tests for

365   each substance for both cell types (3 x 60 x 2 = 360 definitive tests). PC tests were run

366   concurrently with the definitive tests and generally more than one reference substance was

367   tested in conjunction with one PC test plate. Due to test failures, each laboratory performed

368   additional testing to attempt to obtain the three acceptable definitive tests requested for each

369   substance.

370

371   **Table 8-3   Definitive Test and Positive Control (PC) Test Failure Rates**

Test Type	3T3 NRU Test Method				NHK NRU Test Method				Total
	ECBC	FAL	IIVS	Total	ECBC	FAL	IIVS	Total	
Definitive Tests - Acceptable	169	177	176	522	173	175	174	522	1044
Definitive Tests - Total	215	257	225	697	187	256	194	637	1334
% Definitive Tests Failed	21	30	22	25	8	32	10	18	22
PC Tests - Acceptable	66	40	16	122	58	37	20	115	237
PC Tests - Total	74	42	17	133	59	45	20	124	257
% PC Tests Failed	11	5	6	8	2	18	0	7	8
Definitive Tests Failed Only Because PC Tests Failed	14	6	14	34	0	22	0	22	56
% Definitive Tests Failed Only Because PC Tests Failed	7	2	6	5	0	9	0	4	4

372

373   **Table 8-4** illustrates the success rates of the testing for each laboratory and for the combined

374   laboratories.

375

376

376 **Table 8-4   Definitive Test and PC Test Success Rates for 3T3 and NHK NRU Test**
377 **Methods (Combined Total Tests)**

Test Type	ECBC	FAL	IIVS	Total
Acceptable Definitive Tests/ Total Definitive Tests	342/402	352/513	350/419	1044/1334
% Acceptable Definitive Tests	85%	69%	84%	78%
Acceptable PC Tests/Total PC Tests	124/133	77/87	36/37	237/257
% Acceptable PC Tests	93%	89%	97%	92%

378

379 8.3.3   Intralaboratory Reproducibility

380 CV values for each reference substance were determined for each laboratory using the IC_{50}

381 values from the acceptable definitive tests as described in **Section 5.3.1**. **Table 8-5**

382 illustrates the average CV values for the substances tested in each of the phases and for the

383 entire study.

384

385 **Table 8-5   Coefficients of Variation**

Cell Type	Labs	Phases I & II		Phase III		All Phases	
		Number of Reference Substances	Average % CV	Number of Reference Substances	Average % CV	Number of Reference Substances	Average % CV
3T3	ECBC	12	17	57	24	69	23
	FAL	11	28	55	33	66	33
	IIVS	11	20	56	22	68	21
NHK	ECBC	12	24	57	22	69	23
	FAL	12	31	57	45	69	42
	IIVS	12	14	58	14	70	14

386

387 8.3.4   Globally Harmonized System Toxicity Category Predictions

388 Predicted LD_{50} values were compared to the GHS *in vivo* acute oral toxicity categories to

389 determine category match (i.e., accuracy) or toxicity underprediction or overprediction for

390 the reference substances (see **Table 8-6**). Predicted LD_{50} values were determined for the

391 reference substances by using the mean IC_{50} values from the laboratories in the RC

392 regression. The reference GHS *in vivo* acute oral toxicity category presented in **Table 8-6**

393   was the initial LD_{50} value used to select the substances (see **Table 3-1**). The laboratories

394   were generally in agreement with each other in the predictions. Although FAL had the

395   highest error rates and CV values, their predictions of GHS toxicity category using these

396   NRU methods were consistent with the other laboratories. (See **Appendix J** for additional

397   laboratory comparisons for the other *in vitro – in vivo* regressions evaluated in **Section 6**.)

398

399   **Table 8-6   GHS Toxicity Category Predictions by Laboratory[1]**

	Labs	Total Reference Substances	Category Match	Toxicity Overpredicted	Toxicity Underpredicted
3T3	ECBC	69	29%	41%	30%
	FAL	67	28%	43%	28%
	IIVS	69	28%	41%	32%
NHK	ECBC	69	28%	42%	30%
	FAL	69	28%	41%	32%
	IIVS	70	29%	40%	31%

400   [1]GHS-Globally Harmonized System categories of acute oral toxicity with LD_{50} in mg/kg (UN 2003). 3T3
401   and NHK NRU test method IC50 data (geometric mean of within laboratory replicates) used with the RC
402   regression: $\log(LD_{50} \text{ mmol/kg}) = 0.425 \times \log(IC_{50} \text{ mM}) +0.625$.
403

404   **8.4   Availability of Laboratory Notebooks**

405

406   All laboratories maintained laboratory notebooks patterned after a template provided by IIVS

407   and provided copies of them to the SMT (archived at NICEATM) after each phase. The

408   workbooks contained information from all aspects of testing including but not limited to:

409   • environmental conditions

410   • reagent identification

411   • preparation of 96-well plates

412   • preparation of reference substances

413   • treatment of cell cultures

414   • visual observations of cell cultures

415   • NRU assays

416   • data analysis

417

418   **8.5   Summary**

419

420   •   Various determinations of test method and data collection errors consistently

421   showed that FAL had the highest error level; however, the laboratory's GHS

422   acute oral toxicity category predictions were comparable to the other

423   laboratories' results. Data were not adversely affected by general

424   transcriptional errors.

425   •   The laboratories reported no significant deviations from the test method

426   protocols and deviations that did occur during the testing phases were generally

427   quickly acknowledged and addressed by the Study Directors. If a deviation

428   occurred that would affect data (e.g., improper concentration of DMSO

429   solvent), then that Study Director would reject the test, notify the SMT, and

430   perform an additional test. Improper transfer of data to either the EXCEL$^®$ or

431   PRISM$^®$ templates, which would affect the data, were recognized, documented,

432   and rectified by the Study Director and/or the SMT.

433   •   The SMT was diligent in reviewing all data sheets to ensure that data were not

434   inadvertently attributed to the incorrect data summary files and that the correct

435   data were used in all statistical analyses.

436

437   An electronic copy of all data for this validation study can be obtained upon request from

438   NICEATM.

439

439
440
441
442
443
444
445
446
447
448
449
450
451
452
453   ***[This Page Intentionally Left Blank]***
454

27
28

29

30

31

32

33

34

35

36

37

38

39

40

41

42   ***[This Page Intentionally Left Blank]***

43

43 **9.0   OTHER SCIENTIFIC REPORTS AND REVIEWS OF *IN VITRO***

44 **CYTOTOXICITY TEST METHODS AND THE ABILITY OF THESE TEST**

45 **METHODS TO PREDICT ACUTE SYSTEMIC TOXICITY**

46

47 *In vitro* cytotoxicity test methods based on NRU have been evaluated for a number of uses.

48 This section reviews studies relevant to:

49 • the prediction of acute rodent systemic toxicity using *in vitro* NRU cytotoxicity

50   test methods

51 • the use of *in vitro* cytotoxicity test methods to predict starting doses for acute

52   systemic toxicity tests, and

53 • the use of *in vitro* NRU cytotoxicity test methods to predict other *in vivo*

54   endpoints.

55

56 **Section 9.1** discusses *in vitro* studies that evaluated cytotoxicity using NRU for correlation

57 with acute systemic toxicity in rodents and with other *in vivo* endpoints. Also reviewed are

58 studies that have evaluated the use of *in vitro* cytotoxicity results to reduce animal use in

59 acute toxicity testing. **Section 9.2** reviews independent evaluations of the use of *in vitro*

60 cytotoxicity methods to determine starting doses for acute systemic toxicity assays and a

61 validated NRU test method similar to that used in the current study. The conclusions of these

62 reports will be compared to the conclusions reached in this study where possible. **Section**

63 **9.3** reviews studies that have used the *Guidance Document* approach (ICCVAM 2001b),

64 which establishes the current test method performance standard.

65

66 **9.1   Relevant Studies**

67

68 9.1.1   Correlation of *In Vitro* NRU Cytotoxicity Results with Rodent Lethality

69 This section reviews *in vitro* cytotoxicity studies that have used NRU methods to predict

70 rodent lethality. *Italics* identify chemicals tested in the reviewed studies that were also tested

71 in the NICEATM/ECVAM validation study reviewed in this BRD.

72

73

74   *Peloux et al. (1992)*

75   Using several different *in vitro* cytotoxicity test methods with primary rat hepatocytes,

76   Peloux et al. (1992) determined the correlation with rat/mouse intraperitoneal (ip) or

77   intravenous (iv) LD_{50} values for the 25 chemicals tested. The *in vitro* cytotoxicity test

78   methods, which used a 20-hour chemical exposure duration, assessed the following

79   endpoints: NRU; total protein content, lactate dehydrogenase (LDH) release, tetrazolium salt

80   MTT reduction. [NOTE: MTT is metabolized by mitochondrial succinate dehydrogenase of

81   proliferating cells to yield a purple formazan reaction product.] The IC_{50} values obtained

82   using the four endpoints were highly correlated (r = 0.973-0.999) to one another. For the

83   IC_{50}-LD_{50} regressions, Peloux et al. (1992) used the lowest reported LD_{50} value published for

84   rat or mouse studies that administered the test substances acutely using the ip or iv routes.

85   The regressions used units of ln μg/mL for the IC_{50} and ln mg/kg for the LD_{50}. The IC_{50}

86   values obtained using NRU had the highest correlation coefficient, r = 0.877, to the to

87   rat/mouse ip/iv LD_{50} values. The total protein assay yielded r = 0.872, the MTT reduction

88   assay yielded r = 0.808, and the LDH release assay yielded r = 0.789.

89

90   *Fautrel et al. (1993)*

91   Six laboratories tested the cytotoxicity of 31 chemicals in primary rat hepatocytes using a 24-

92   hour exposure followed by measuring NRU. The investigators performed linear regression

93   analyses for the prediction of rat iv, ip, and oral LD_{50} values by the NRU IC_{50} values. The

94   regressions by the various *in vivo* administration routes did not use the same chemicals since

95   LD_{50} values for all the routes were not available for all the tested chemicals. Oral , iv, and ip

96   LD_{50} values were available for 27, 24, and 18 chemicals, respectively. IC_{50} values were

97   obtained for 15, 14, and 11 of the chemicals, respectively. The units used for correlation

98   were ln μg/mL for the IC_{50} and ln mg/kg for the LD_{50}. While the regression for the iv data

99   was statistically significant (r = 0.88, n=11), the ip (r=0.48, n=14) and oral regressions

100   (r=0.17, n=15) were not. The fact that the parenteral LD_{50} values correspond more closely

101   with *in vitro* cytotoxicity data than do the oral LD_{50} was thought to be due to the fact that

102   there are fewer kinetic variables (i.e., absorption, distribution, etc.) to consider for iv

103   administration. The authors concluded that the hepatocyte cultures were useful in screening

104   chemical classes with high bioavailability.

105   *Roguet et al. (1993)*

106   Roguet et al. (1993) tested the cytotoxicity of 28 MEIC chemicals in primary rat hepatocytes

107   exposed to the chemicals for 21 hours, followed by measuring NRU. A correlation of the

108   NRU IC_{50} values to LD_{50} values obtained from the unpublished data of B Ekwall et al.

109   yielded a statistically significant linear correlation (p < 0.001) with r = 0.80. [NOTE: The

110   LD_{50} values subsequently published by Ekwall et al. (1998) were from the 1997 edition of

111   RTECS[®].] The correlation used molar units for the *in vivo* and *in vitro* data. Roguet et al.

112   (1993) reported that the toxicities of thioridazine, malathion, and *copper sulfate* were

113   overestimated and the toxicity of *potassium cyanide* was underestimated, but their criterion

114   for over/under estimation was not provided. The toxicity of *potassium cyanide* was also

115   underpredicted (see **Appendix L-2**) when using the Registry of Cytotoxicity (RC) rat only

116   weight regression (i.e., log LD_{50} = 0.372 log IC_{50} + 2.024) prediction of GHS toxicity

117   categories by the NICEATM/ECVAM 3T3 and NHK NRU test methods. The RC is a

118   database of acute oral LD_{50} values for rats and mice obtained from RTECS[®] and IC_{50} values

119   from *in vitro* cytotoxicity assays using multiple cell lines and cytotoxicity endpoints for

120   chemicals with known molecular weights (Halle 1998).

121

122

123   *Rasmussen (1999)*

124   Twenty MEIC chemicals were tested for cytotoxicity in 3T3 cells using NRU with and

125   without the addition of Arochlor-induced rat liver microsomes (S9 mix). The chemical

126   exposure duration was 24 hours. Similar to the present validation study, Rasmussen (1999)

127   was unable to attain cytotoxicity with *xylene*, although it was dissolved in ethanol instead of

128   DMSO. In the presence of S9, the cytotoxicities of malathion, 2,4-dichlorophenoxyacetic

129   acid, *propranolol*, thioridazine, *lithium* sulfate, *copper sulfate*, and *thallium sulfate* were

130   significantly decreased (p<0.05) while the cytotoxicities of *1,1,1-trichloroethane, phenol,*

131   *nicotine,* and *paraquat* were significantly increased (p<0.05).

132

133   The toxicities of *nicotine, thallium sulfate,* and *paraquat* were also underpredicted in the

134   NICEATM/ECVAM validation study (see **Appendix L-2**) when using the RC rat only

135   weight regression (i.e., log LD_{50} = 0.372 log IC_{50} + 2.024) prediction of GHS toxicity

136   categories by 3T3 and NHK NRU test methods.

137

138   Although both IC_{20} and IC_{50} values were determined in the Rasmussen (1999) study, only the

139   IC_{20} values were used for correlations with rat acute oral LD_{50} values from RTECS[®]. Even

140   though the units of the LD_{50} values were not reported, the correlations are assumed to be in

141   molar units since the IC_{20} and IC_{50} values were reported in μM units. Significant linear

142   correlations (p< 0.001) for IC_{20} and LD_{50} values were obtained with and without

143   microsomes. The correlation was slightly higher with microsomal activation (r = 0.72 vs.

144   0.68 for oral and 0.82 vs. 0.78 for ip).

145

146   Although the presence of S9 increased the cytotoxicity of some chemicals, it decreased the

147   toxicity of others, and yielded only a small improvement in the correlation to *in vivo* data.

148

149   *Creppy et al. 2004*

150   Creppy et al. (2004) used a 48-hour NRU assay to determine the cytotoxicity of ochratoxin A

151   (OTA) and fumonisin B1 (FB1) on C6 glioma (rat brain), Caco-2 (human intestinal), and

152   Vero (green monkey kidney) cells. The IC_{50} determined in the NRU assay was used in the

153   RC regression to predict the acute oral LD_{50}. The predicted LD_{50} using the C6 glioma cells

154   was similar to mouse LD_{50} values (data generated from four *in vivo* studies), but the LD_{50}

155   values predicted by the other cell lines were about 50 times greater than that predicted by the

156   C6 glioma cells. The authors found the relative insensitivity of the Vero cells surprising

157   since OTA was known to be a kidney toxin. There were no LD_{50} values with which to

158   compare the predicted LD_{50} of FB1.

159

160   9.1.2   Use of Cytotoxicity Data to Reduce the Use of Animals in Acute Toxicity Testing

161   *Halle et al. (1997): Animal Savings Using Cytotoxicity Data with the ATC*

162   This study predicted the animal savings that would be produced by using IC_{50} data from

163   cytotoxicity tests in the RC regression to determine a starting dose for ATC testing. No

164   cytotoxicity testing was performed for this study. The authors used the IC_{50x} data from the

165   RC and the RC regression to predict the LD_{50} for the 347 RC chemicals. At the time of the

166   Hall et al. (1997) study, the ATC (1996 version from OECD) was designed to classify

167   chemicals using the three classes of acute oral toxicity and an unclassified group defined by

168   the acute oral toxicity classification system of the European Union (EU) (see **Table 9-1**).

169   Thus, the fixed doses for the ATC testing were 25, 200, and 2000 mg/kg.

170

171   **Table 9-1   EU[1] Classes of Acute Oral Toxicity**

172

Category	LD_{50} (mg/kg)
1	$LD_{50} \leq 25$
2	$25 < LD_{50} \leq 200$
3	$200 < LD_{50} \leq 2000$
Unclassified	$LD_{50} > 2000$

173   [1]Anon (1993)

174

175   Halle et al. (1997) used the RC predicted LD_{50} for the 347 RC chemicals as a starting point to

176   estimate the number of ATC dose steps (and animals required) that would be needed to

177   classify the chemicals in the same EU category associated with *in vivo* LD_{50} (i.e., oral rat or

178   mouse values from RTECS®). The method required the simulated ATC testing for each

179   chemical to start at the nearest fixed ATC dose to the LD_{50} predicted by the RC. The

180   outcome of the simulated testing of three animals per fixed dose was determined by the *in*

181   *vivo* LD_{50}. If the test dose was lower than the *in vivo* LD_{50}, animals were assumed to live

182   and, if the test dose was higher than the LD_{50}, the animals were assumed to die. Testing of

183   the chemical would proceed with higher (when the animals lived) or lower fixed doses (when

184   the animals died) until the chemical was placed into the EU toxicity category indicated by the

185   *in vivo* rodent oral LD_{50}.

186

187   The method of Halle et al. (1997) can be illustrated with digoxin, which has an *in vivo* rodent

188   LD_{50} of 18 mg/kg (from RTECS®) and an RC predicted LD_{50} of 414 mg/kg. Simulated ATC

189   testing would start at the nearest fixed dose, 200 mg/kg, to the RC predicted LD_{50} of 414

190   mg/kg. The three animals tested would die, and three more animals would be tested at 25

191   mg/kg. The animals tested at 25 mg/kg would die and digoxin would be classified in

192   category 1 for $LD_{50} \leq 25$ mg/kg. Thus, classification of digoxin required six animals.

193

194   Using such simulations of ATC testing for the 347 RC substances, Halle et al. (1997)

195   estimated a total of 2139 test animals would be used:

196   • 328 substances would require testing with two doses with three test animals

197   each

198   • 19 substances would require testing with three doses with three animals each

199   Halle et al. cited (from Schlede et al. 1995) that the average number of animals required to

200   classify chemicals using the ATC method was 9.11. Using this average, ATC testing of the

201   347 RC chemicals would require 3161 animals. Thus, there would be a 32% reduction in the

202   number of test animals used (compared to the average) when the RC LD_{50} prediction was

203   used in conjunction with the 1996 version of the ATC method (Halle et al. 1997).

204

205   Depending on the regression evaluated, the average animal savings for the ATC predicted in

206   the NICEATM/ECVAM validation study at dose-response slopes of 2 and 8.3 were 8.0 –

207   14.8% (0.85-1.56 animals) for the 3T3 NRU and 8.9 -13.5% (0.97-1.43 animals) for the

208   NHK NRU for the 72 reference substances tested (see **Section 10.3**). This is quite a bit

209   lower than the average savings of 32% calculated by Halle et al. (1997). However, there

210   were a number of differences between the evaluation performed by the Halle et al. (1997)

211   and the NICEATM/ECVAM study that contribute to the difference in calculated animal

212   savings:

213   • the NICEATM/ECVAM study used *in vitro* cytotoxicity data to estimate

214   starting doses (using several regressions based on the RC data)

215   • the chemicals tested in the NICEATM/ECVAM study were different from the

216   RC chemicals (i.e., the 58 RC chemicals tested had a regression significantly

217   different from the RC regression [see **Section 6.1.2**])

218   • the NICEATM/ECVAM study used computer simulations of ATC testing,

219   which incorporated assumptions about mortality distributions, to determine

220   animals used whereas Halle et al. (1997) used simplified assumptions (i.e.,

221   animals live when test dose is less than LD_{50} and die when test dose is greater

222   than LD_{50})

223   • the NICEATM/ECVAM study determined animal savings by comparing animal

224   use with starting doses determined by the *in vitro* data to animals used at the

225   default starting dose of 300 mg/kg. Halle et al. (1997) used the average animal

226   use for the ATC as a comparison to animal use with simulated testing..

227   •   the NICEATM/ECVAM study used the GHS acute toxicity categories for

228   classification whereas Halle et al. (1997) used the EU toxicity classification

229   scheme, which had fewer toxicity categories (i.e., accuracy of category

230   prediction by any method would be higher with fewer categories).

231

232   *Spielmann et al. (1999): Animal Savings Using Cytotoxicity Data with the UDP*

233   Spielmann et al. (1999) recommended an *in vitro* cytotoxicity procedure for reducing the

234   number of animals used in acute toxicity tests. The procedure used *in vitro* cytotoxicity data

235   as a range finding test for the *in vivo* toxicity test.

236

237   The authors identified nine chemicals in common when comparing the RC database to an

238   evaluation of acute toxicity methods by Lipnick et al. (1995). Spielmann et al. (1999)

239   compared the LD_{50} values from Lipnick et al. (1995) to LD_{50} predictions calculated when

240   using the RC IC_{50} values in the RC regression formula. For seven of the nine chemicals, the

241   LD_{50} prediction was within an order of magnitude of the conventional LD_{50} (OECD 1987)

242   used in Lipnick et al. (1995). Spielmann et al. (1999) concluded that the RC provides an

243   adequate prediction of LD_{50} and that cytotoxicity data could be used to predict starting doses

244   for the UDP. If an IC_{50} is available for a particular chemical, the authors recommend using

245   the IC_{50}, with the RC regression, to calculate a starting dose (i.e., estimated LD_{50}) for the

246   UDP, FDP, or ATC method.

247

248   If no IC_{50} is available for a particular chemical, Spielmann et al. (1997) recommended

249   determining cytotoxicity using a standard cell line and specific endpoint of cytotoxicity (e.g.,

250   NRU, total protein, MTT reduction, etc.). To show that the *in vitro* cytotoxicity test methods

251   provide results that are consistent with the RC, Spielmann et al. (1999) recommended testing

252   10-20 RC chemicals. The IC_{50} data are then used to calculate a new regression, which is then

253   compared to the RC regression. If the new regression fits into the acceptance interval (± log

254   5 of the fitted regression line) of the RC regression line, the RC regression is used to predict

255   starting doses for the UDP. If the new regression is parallel to the RC regression, but outside

256   the ± log 5 acceptance interval, Spielmann et al. (1999) recommended using the new

257   regression line for the prediction of the starting dose.

258

259   Spielmann et al. (1999) contends that the RC provides a sufficient prediction of LD_{50} values

260   from IC_{50} values for chemicals that do not require metabolic activation and are not usually

261   toxic (i.e., $LD_{50} > 200$ mg/kg), such as industrial chemicals. The authors acknowledge that

262   the fit of chemicals with $LD_{50} < 200$ mg/kg to the RC regression is not good and attribute the

263   poor fit of these chemicals to the fact that they require metabolic activation for toxicity.

264   They indicated that the prediction of starting doses using cytotoxicity data can be applied to

265   the UDP and ATC methods, but not to the FDP since dosing is not sequential (this

266   contradicted a claim made earlier in the paper that the approach could be used with the FDP).

267   They did not estimate the number of animals that might be saved with this approach, but they

268   did recommend that the approach be validated experimentally using several established cell

269   lines with a limited number of representative chemicals from the RC.

270

271   *EPA (2004): U.S. EPA HPV Challenge Program Submission*

272   PPG Industries, Inc. is the manufacturer of Propanoic acid, 2-hydroxy-, compd. with 3-[2-

273   (dimethylamino)ethyl] 1-(2-ethylhexyl) (4-methyl-1,3-phenylene)bis[carbamate] (1:1) [CAS

274   No. 68227-46-3] and is the sponsor of this compound for the EPA HPV Chemical Challenge

275   Program. The compound is an isolated intermediate and subsequently is used to produce a

276   resin component of paint products. PPG provided the following data on the compound in

277   their submission (http://www.**epa**.gov/chemrtk/prop2hyd/c13863rt3.pdf) to the EPA:

278   physical-chemical, environmental fate and pathway, ecotoxicity, and toxicological. The

279   acute mammalian toxicological data were generated using *in vitro* and *in vivo* test methods.

280

281   An *in vitro* NRU cytotoxicity test with BALB/c 3T3 cells was conducted to estimate a

282   starting dose for the *in vivo* acute oral toxicity study using the UDP (OECD 2001a) (see

283   **Appendix M** for the OECD UDP test guideline). Use of *in vitro* methods was intended to

284   minimize the number of animals used for *in vivo* testing. The estimated LD_{50} of the

285   compound determined by the NRU assay was 489 mg/kg. The starting dose for the UDP

286   study was set at 175 mg/kg, the first default dose below the estimated LD_{50} value. The

287   starting dose of 175 mg/kg is also the default starting dose, which is used when no

288   information (on which to base a starting dose) is available . A total of fifteen female rats

289   received the compound at 175, 550, or 2000 mg/kg. Five of nine rats treated at 2000 mg/kg

290   died prematurely on Days 2 and 3. At 2000 mg/kg, 2/4 surviving animals had lost up to 25%

291   of their Day 1 body weights by Day 15. The LD_{50} for the compound was estimated to be

292   2000 mg/kg with a 95% confidence interval of 1123-5700 mg/kg.   Thus, the *in vitro* NRU

293   cytotoxicity test method overpredicted the toxicity of the compound by estimating a lower

294   LD_{50} value than that determined in the acute oral toxicity UDP study. The report authors felt

295   that a greater than predicted number of animals was used for UDP testing since the LD_{50}

296   estimated by the 3T3 NRU assay, 489 mg/kg, and, consequently, the starting dose, was much

297   lower than the *in vivo* LD_{50} of 2000 mg/kg. However, since the UDP started with the default

298   starting dose of 175 mg/kg, the claim that more animals were used is unfounded, since

299   animal use with the default starting dose is the baseline with which animal use should be

300   compared.

301

302   9.1.3   Other Evaluations of 3T3 or NHK NRU Test Methods

303   This section briefly reviews studies that have evaluated NRU methods for purposes other

304   than the prediction of starting doses for acute oral systemic toxicity assays. NRU test

305   methods using either 3T3 or NHK cells have been evaluated for use as alternatives to the

306   Draize eye irritation test and to predict acute lethality in humans. Except for the 3T3 NRU

307   phototoxicity assay, NRU methods have neither been scientifically validated by an

308   independent review for any of these purposes nor accepted for regulatory use. The use of the

309   3T3 NRU test method to determine phototoxic potential is addressed in **Section 9.2** since it

310   has been validated.

311

312   Based on the method of Borenfreund and Puerner (1985), the *in vitro* NRU test method

313   protocols evaluated in the reviewed studies are similar to those evaluated in the current study.

314   The major difference is that most use a 24-hour chemical exposure duration for the 3T3

315   assay, while the current 3T3 validation study used a 48-hour exposure duration. The major

316   difference between the NHK NRU test method protocols used in these studies and the

317   protocol used in the NICEATM/ECVAM study is the change of medium with test chemical

318   application used in the validation study presented in this BRD.

319

320   *Draize eye irritation*

321   *Triglia et al. (1989)*

322   Four laboratories collaborated in an interlaboratory validation study to test the NHK NRU

323   assay from Clonetics® Corporation. The evaluation included intra- and inter-laboratory

324   reproducibility and the ability to predict *in vivo* ocular irritancy. Each laboratory tested 11

325   surfactant-based test agents and compared the IC_{50} values to available *in vivo* Draize ocular

326   irritancy scores.

327

328   The authors determined the following performance characteristics when comparing the *in

329   vitro* and *in vivo* data:

330   • specificity (percentage of non irritants detected) = 93%

331   • sensitivity (percentage of true irritants detected) = 80%

332   • predictive values (probability that an unknown agent will be properly classified)

333   – positive predictive value = 90%

334   – negative predictive value = 87%

335

336   The authors concluded there was excellent correlation among the laboratories and good

337   correlation between the *in vitro* NR_{50} values (concentration at 50% reduction of NRU

338   compared to controls) and the Draize data (Spearman Rank correlation coefficients between

339   *in vivo* and *in vitro* data for the laboratories ranged from 0.67-0.76). The authors also

340   concluded, however, that the NRU assay could not replace the Draize test but may be an

341   effective screening tool for use in a battery of *in vitro* alternatives.

342

343   *Sina et al. (1995)*

344   Sina et al (1995) evaluated the NHK NRU test method along with six other *in vitro* test

345   methods to evaluate whether they could be used as complimentary tests in a battery

346   approach. The NRU data correlated poorly with Draize scores for the 33 pharmaceutical

347   intermediates that were tested. The Spearman correlation coefficient for the NR_{50} and

348   maximum average Draize score (MAS) was -0.10 and the Pearson correlation coefficient was

349   -0.04.

350

351   *Brantom et al. (1997)*

352   This study examined the potential of 10 alternative methods to predict the eye irritation

353   potential of cosmetic ingredients. Four laboratories tested 55 coded substances (23

354   ingredients and 32 formulations) with the 3T3 NRU test method and used the resulting IC_{50}

355   to predict modified maximum average scores (MMAS) for the Draize test.

356

357   An endpoint in µg/mL was generated for each test by interpolation from a plot of percentage

358   cell survival versus the test substance concentration. A prediction model (PM) was

359   developed from data of 30 single ingredients (29 surfactants and one chemical not classified

360   by the authors) to equate the IC_{50} value to an MMAS.

361

362   The interlaboratory CV for the NR_{50} values was $37.3 \pm 29.8\%$ (7.5 ± 6.8 log transformed).

363   No mean IC_{50} value for a single laboratory differed by an order of magnitude from the mean

364   of all the laboratories for each chemical, which the authors interpreted as "no significant

365   outliers". Correlations of NRU predicted MMAS scores with *in vivo* MMAS scores yielded

366   Pearson's $r = 0.25 – 0.32$ (for the four laboratories).

367

368   Although the authors concluded the reproducibility was good, the data did not accurately

369   predict the MMAS (i.e., low r values for *in vitro/in vivo* correlations; underpredicted irritants,

370   overpredicted non-irritants). However, the authors concluded that the 3T3 NRU test method

371   had wide applicability to test 51/55 coded substances according to the limitations in the

372   prediction model (four substances outside of the 95% confidence intervals), but that it was

373   not a stand-alone replacement for the Draize test across the entire irritation scale. None of

374   the substances tested were identified by the authors.

375

376   *Harbell et al. (1997)*

377   This paper reported the results of the evaluation of 12 *in vitro* cytotoxicity assays to predict

378   ocular irritation. Data were voluntarily submitted to the US Interagency Regulatory

379   Alternatives Group (IRAG), composed of members from CPSC, EPA, and FDA. The NHK

380   NRU test method was one of the tests evaluated by six laboratories testing surfactants and

381   surfactant-containing formulations (the 3T3 NRU was not tested). Two laboratories

382   submitted results for the same test substances, but the other four laboratories submitted data

383   for various sets of chemicals and formulations.

384

385   The correlation of results from the two laboratories testing the same substances was r = 0.99.

386   Correlations between the NR_{50} data and *in vivo* maximum average score (MAS) ranged from

387   -0.92 to -0.54. The IRAG concluded that the assay was suitable as a screening and adjunct

388   assay to assess eye irritation over the range of toxicities found in personal care and household

389   products. IRAG recommends that its use be limited to water-soluble materials. Although the

390   method was evaluated for surfactants, IRAG recommended that the evaluation continue for

391   its performance in predicting eye irritation for various product classes (e.g., fabric softeners,

392   shampoos). IRAG also recommended that physical form be considered since toxicity of the

393   solution (*in vitro*) does not necessarily predict toxicity of the solid (*in vivo*).

394

395   *Predicting human lethal blood concentrations*

396   *Seibert et al. (1992)*

397   The aim of this single laboratory study was to evaluate various aspects of cellular toxicity in

398   four *in vitro* test systems for relevance and reliability to acute systemic toxicity, in particular,

399   human lethal blood concentrations. The 3T3 NRU test method was one of four methods

400   evaluated with 10 MEIC chemicals.

401

402   The authors stated that final conclusions on the relevance of the *in vitro* systems could not be

403   determined when compared to the *in vivo* data. The variations in lethal blood concentrations

404   are unknown and make it difficult to define limits for over/underprediction of *in vivo* toxicity

405   using experimental models. In addition, the ability of *in vitro* toxicity to predict *in vivo*

406   toxicity may strongly depend on toxicokinetic factors.

407

408   **9.2   Independent Scientific Reviews**

409

410   This section **(9.2)** covers independent scientific reviews of the use of *in vitro* cytotoxicity

411   methods for the prediction of acute oral toxicity and reduction of animal use. The

412   conclusions of these reports are compared to the conclusions of the current study. Also

413   discussed is the 3T3 NRU phototoxicity test method that has been validated by ECVAM.

414

415   9.2.1   Use of *In Vitro* Cytotoxicity Data for Estimation of Starting Doses for Acute Oral

416   Toxicity Testing

417   *ICCVAM (2001a): Estimation of Animal Savings Using Cytotoxicity Data with the ATC*

418   Participants at Workshop 2000 examined the influence of starting dose on animal usage for

419   the ATC method (ICCVAM 2001a, section 2.2.3, pp.12-14). No testing was performed at

420   the Workshop. The participants made inferences from the 1996 version of the ATC method

421   that was based on the EU hazard (i.e., toxicity) classification system in **Table 9-1**. The fixed

422   doses for testing were 25, 200, and 2000 mg/kg. Normally, classification of a substance

423   requires testing three animals in two to four dosing steps (i.e., six to 12 animals). With

424   increasing distance between the true toxicity class and the starting dose, the number of

425   dosing steps increases. They estimated that one to three dosing steps could be avoided if the

426   optimum starting dose could be predicted by *in vitro* cytotoxicity (i.e., three to nine animals

427   saved).

428

429   The savings of one to three dosing steps was predicted under ideal conditions. The

430   Workshop 2000 report (ICCVAM 2001a) provides a biometrical analysis at a dose-mortality

431   slope of 2 by W. Diener that shows that the largest animal savings occur for chemicals with

432   very high and very low toxicity. Three animals are needed to classify a chemical in the < 25

433   mg/kg class if the true LD_{50} is 1 mg/kg and 25 mg/kg is the starting dose, but six animals are

434   needed if the test starts from the default starting dose of 200 mg/kg (i.e., animal savings =

435   33%). For a chemical with a true LD_{50} of 10000 mg/kg, 11.3 animals on average are needed

436   using the default starting dose, but only 7.7 animals are needed at the 2000 mg/kg starting

437   dose (i.e., animal savings = 31%). For chemicals with a true LD_{50} of 2000 mg/kg, no

438   animals are saved by starting at the 2000 mg/kg dose (compared to starting at the default

439   starting dose of 200 mg/kg).

440

441   Although these analyses were performed assuming the 1996 ATC method used starting doses

442   of 25, 200, 2000 mg/kg, Workshop 2000 participants expected that animal savings that would

443   be produced by improving the starting dose would not be significantly different for the

444   current ATC method that uses GHS doses of 5, 50, 300, and 2000 mg/kg (or up to 5000

445   mg/kg) (OECD 2001c; see **Appendix M** for the current ATC test guideline).

446

447   Beyond presenting the biometrical analysis by W. Diener, Workshop 2000 participants did

448   not predict the animal savings when *in vitro* cytotoxicity methods are used to estimate

449   starting doses for the ATC.

450

451   The NICEATM/ECVAM study yielded a pattern of animal savings for the ATC that was

452   similar to those discussed at the 2000 Workshop (i.e., animal savings were greater for

453   chemicals with lower or higher LD_{50} than the default starting dose; see **Section 10.3**).

454   Depending on the regression evaluated, the average animal savings (for the 72 reference

455   substances tested) predicted by the NICEATM/ECVAM validation study at a dose-response

456   slope of 2 was:

457   •   12.8-17.1 % (1.22-1.63 animals) for the 3T3 NRU and 7.6-13.0% (0.72-1.23

458   animals) for the NHK NRU for chemicals in the $LD_{50} \leq 5$ mg/kg category

459   •   12.1-16.6 % (1.45-1.98 animals) for the 3T3 NRU and 18.9-23.9% (2.26-2.86

460   animals) for the NHK NRU for chemicals in the $5 < LD_{50} \leq 50$ mg/kg category

461   •   3.6-4.3 % (0.39-0.47 animals) for the 3T3 NRU and 2.1-2.8% (0.23-0.30

462   animals) for the NHK NRU for chemicals in the $50 < LD_{50} \leq 300$ mg/kg

463   category

464   •   -2.8- -0.2% (-0.24 - -0.02 animals) for the 3T3 NRU and -1%-0.8% (-0.10-0.02

465   animals) for the NHK NRU for chemicals in the $300 < LD_{50} \leq 2000$ mg/kg

466   category

467   •   1.4-14% (0.16–1.67 animals) for the 3T3 NRU and 3.4%-11.0% (0.38-1.23

468   animals) for the NHK NRU for chemicals in the $2000 < LD_{50} \leq 5000$ mg/kg

469   category

470   •   16.2-31.1% (1.92-3.70 animals) for the 3T3 NRU and 14.2-29.2% (1.69-3.47

471   animals) for the NHK NRU for chemicals with $LD_{50} > 5000$ mg/kg

472

473   The major differences between the evaluation reviewed by the Workshop 2000 participants

474   and the NICEATM/ECVAM study were:

475   • the NICEATM/ECVAM study used *in vitro* cytotoxicity data to estimate

476   starting doses (using several regressions based on the RC data), whereas the

477   Workshop 2000 participants used the fixed ATC doses as starting doses

478   • the NICEATM/ECVAM study used computer simulations of ATC testing for

479   individual chemicals whereas Workshop 2000 participants used an evaluation

480   that provided animal use based on fixed *in vivo* LD_{50} values and the fixed ATC

481   doses

482   • the NICEATM/ECVAM study used the GHS acute toxicity categories for

483   classification whereas the Workshop participants used the EU classification

484   scheme which had fewer toxicity categories (i.e., accuracy of category

485   prediction by any method would be higher with fewer categories)

486

487   *ICCVAM (2001a): Estimation of Animal Savings Using Cytotoxicity Data with the UDP*

488   Workshop 2000 participants examined the effect of starting dose on animal usage in the UDP

489   assay by making inferences from computer simulations of animal use shown in the UDP peer

490   review BRD (ICCVAM 2000). Using the rule that requires testing to stop when four animals

491   have been tested after the first reversal (and no other stopping rules), animal use is relatively

492   insensitive to the slope of the dose-mortality curve. The number of animals required when

493   the starting dose equals the true LD_{50} is approximately six. When the starting dose is 1/100

494   times the true LD_{50}, however, approximately nine animals are required. Thus, animal use is

495   30% less when the starting dose is the true LD_{50} compared to a starting dose of 1/100 times

496   the true LD_{50} (ICCVAM 2001a, section 2.2.4, pg. 16). When the UDP testing stops based on

497   the likelihood-ratio stopping rule, animal use depends heavily on the slope of the dose-

498   mortality curve. Workshop 2000 participants estimated that 25-40% animals would be saved

499   when the starting dose is equal to the true LD_{50} compared to a starting dose of 1/100 times

500   the true LD_{50}.

501

502   At a slope of 0.5, on average 12.4 animals were predicted to be used when the starting dose is

503   1/100 times the true LD_{50}, but use of an average of 8.7 animals was predicted when the

504   starting dose equals the true LD_{50} (30% reduction). At a slope of 8.3, an average of 11

505   animals were predicted to be used when the starting dose is 1/100 times the true LD_{50}, but an

506   average of only six animals are used when the starting dose equals the true LD_{50} (46%

507   reduction).

508

509   The animal savings predicted by Workshop 2000 participants were 25-40% based on starting

510   at the true LD_{50} in comparison to starting at a dose 1/100 times the LD_{50} as the starting dose.

511

512   Depending on the regression evaluated, the average animal savings predicted in the

513   NICEATM/ECVAM validation study at dose-response slopes of 2 and 8.3 were 6.6 - 13.0%

514   (0.63-1.25 animals) for the 3T3 NRU and 6.7 -12.9% (0.64-1.23 animals) for the NHK NRU

515   for the 72 reference substances tested (see **Section 10.2**). When calculated for the chemicals

516   in each GHS toxicity category, the highest average animal savings at a dose-response slope

517   of 2 was for chemicals in the $2000 < LD_{50} \le 5000$ mg/kg category. Animal savings was

518   predicted to be 22.6-26.2% for the 3T3 NRU and 21.0-26.0% for the NHK NRU, depending

519   upon the regression used. The highest average animal savings at a dose-response slope of 8.3

520   was for chemicals in the $LD_{50} > 5000$ mg/kg group. Animal savings was predicted to be

521   26.8-32.0% for the 3T3 NRU and 23.2-30.6% for the NHK NRU, depending upon the

522   regression used.. The major differences between the evaluation performed by the Workshop

523   2000 participants and the NICEATM/ECVAM study were that:

524   • the comparison default starting dose used for the NICEATM/ECVAM

525   simulations was 175 mg/kg, rather than 1/100 times the true LD_{50} assumed by

526   the Workshop 2000 participants (see **Section 10.2**).

527   • the NRU IC_{50} was used in various regressions of *in vitro* data against *in vivo*

528   data to estimate starting doses. This estimation was not always close to the true

529   LD_{50}, which was used by the Workshop 2000 participants. For example, the

530   starting doses predicted by the NICEATM/ECVAM study for phenylthiourea

531   were approximately 800 mg/kg by the 3T3 NRU and approximately 1250 mg/kg

532   by the NHK NRU (see **Appendix N**). The true in vivo LD_{50} for phenylthiourea

533   is 3 mg/kg. Workshop 2000 participants used a best case scenario when they

534   assumed that *in vitro* cytotoxicity predicted exactly the true LD_{50}.

535

536   9.2.2   Validation of 3T3 NRU for Phototoxicity

537   An NRU assay using 3T3 cells was validated by ECVAM and accepted for regulatory use to

538   detect the phototoxic potential of test substances. The 3T3 NRU test for phototoxicity

539   requires a 60-minute exposure to test chemicals, a 50-minute exposure to ultraviolet (UVA,

540   315-400 nm) light, and then removal of test chemical (Spielmann et al. 1998). After

541   incubation for another 24 hours in fresh medium, NR medium is added and NRU is measured

542   after a 3-hour incubation. Phototoxic potential is assessed by comparing the differences in

543   cytotoxicity between negative control test plates that have not been exposed to UVA and test

544   plates exposed to UVA.

545

546   Two different models, the Photoinhibition Factor (PIF) and the Mean Photo Effect (MPE),

547   for the prediction of *in vivo* phototoxic potential were validated. The accuracy of the models

548   for classifying the phototoxic potential of the 30 chemicals tested in nine laboratories was

549   88% for the PIF and 92% for the MPE when compared with *in vivo* classifications.

550   Interlaboratory variability for classification (i.e., phototoxic vs. non-phototoxic) was assessed

551   using a bootstrapping approach. For each chemical, classification based on a single

552   experiment was compared to classification based on the mean PIF or mean MPE.

553   Interlaboratory variability for classification was 0-18.8% for PIF and 0-20% for MPE.

554

555   The ECVAM Scientific Advisory Committee confirmed the scientific validity of the method

556   in 1997 (ECVAM 1997) and its regulatory acceptance was noted in Annex V of Council

557   Directive 67/548/EEC part B.41 on phototoxicity in 2000. An OECD test guideline, 432,

558   was finalized in 2004 (OECD 2004). The test results from the 3T3 NRU phototoxicity test

559   are used in a tiered testing approach to determine the phototoxic potential of test substances.

560

561   Performance of the 3T3 NRU phototoxicity assay could not be compared to the performance

562   of the 3T3 NRU test method used in this validation study since different classification

563   schemes were used (i.e., a two category classification for the phototoxicity vs. a six class

564   scheme for acute oral toxicity). Measurements of interlaboratory variability also used

565   different techniques and were not comparable to those used for the NICEATM/ECVAM

566   study.

567

568   *NHK NRU Phototoxicity Assay*

569   FAL participated in the European Union/European Cosmetic, Toiletry and Perfumery

570   Association (EU/COLIPA) study (30 chemicals using NHK and 3T3 cells) and the

571   ECVAM/COLIPA study (20 chemicals using only NHK cells) (Clothier et al. 1999). The

572   authors showed that the NHK NRU test method could also be used to predict phototoxic

573   potential. The accuracy for predicting *in vivo* results was similar to that of the 3T3 NRU

574   phototoxicity test (see **Table 9-2**). The NHK NRU phototoxicity test uses the same chemical

575   exposure duration (approximately 2 hours) as the 3T3 NRU phototoxicity test, but the

576   duration of culture after UV exposure is 72 hours rather than 24 hours. NRU was measured

577   after a 45-minute incubation with NR.

578

579   **Table 9-2   Correct Predictions of *In Vivo* Phototoxicants by the NHK NRU**

580   **   Phototoxicity Assay**

Study	3T3 NRU Photoxicity Test Method	NHK NRU Photoxicity Test Method
EU/COLIPA (Spielmann et al. 1998)	29/30 (97%)[1]	28/30 (93%)[1]
ECVAM/COLIPA	NA	18/20 (90%)[1] 19/20 (95%)[2]
Combined Study Data	45/45 (100%)[2]	44/45 (98%)[2]

581   [1]Mean Photo Effect prediction model
582   [2]Photoinhibition Factor prediction model
583   NA – not available
584

585   Although the NHK NRU phototoxicity test method achieved good concordance with *in vivo*

586   phototoxicity, it has not been validated for regulatory use.

587

588   **9.3   Studies Using *In Vitro* Cytotoxicity Test Methods with Established**

589   **   Performance Standards**

590

591   The *Guidance Document* method of evaluating basal cytotoxicity assays for use in predicting

592   starting doses for acute oral toxicity assays provides the existing performance standards for

593   the 3T3 and NRU test methods (ICCVAM 2001b).

594

595   9.3.1   *Guidance Document* (ICCVAM 2001b)

596   In addition to instructions for evaluating basal cytotoxicity methods for use in predicting

597   starting doses for acute systemic toxicity assays, the *Guidance Document* provided results

598   from testing 11 reference chemicals (ICCVAM 2001b). The 11 reference chemicals were

599   tested with the 3T3 and NHK NRU test method protocols recommended in the *Guidance*

600   *Document*. The 11 chemicals were chosen from the RC database so as to have a close fit to

601   the RC IC_{50} –LD_{50} regression and to cover a wide range of cytotoxicity. The major

602   differences in the *Guidance Document* protocols and the protocols used in this study are the

603   reduced NR concentrations (from 50 μg/mL to 25 μg/mL in the 3T3 assay and to 33 μg/mL

604   in the NHK assay), the increased chemical exposure duration for the 3T3 test method (from

605   24 to 48 hours), and the lack of a refeeding step for the NHK test method just prior to

606   chemical application (see **Section 2.2** for further detail). Nevertheless, the *Guidance*

607   *Document* shows the similarity of the results for the 11 chemicals in both the 3T3 and NHK

608   NRU test methods to the RC data. The regressions were:

609   • log (LD_{50}) = 0.506 (log IC_{50}) + 0.475 (R^2 = 0.985) for the 3T3 NRU

610   • log (LD_{50}) = 0.498 (log IC_{50}) + 0.551 (R^2=0.936) for the NHK NRU, and

611   • log (LD_{50}) = 0.435 (log IC_{50}) + 0.625 for the RC.

612

613   The 3T3 and NHK NRU regressions were graphed on the RC regression (347 chemicals) to

614   show that the regression lines as well as all 11 chemical data points were within the

615   acceptance interval (± 0.5 log around the regression) of the RC regression (see **Appendix D-**

616   **1**, *Guidance Document*, **Figures 3** and **4, pg.13**).

617

618   9.3.2   King and Jones (2003)

619   This study also tested the 11 chemicals recommend in the *Guidance Document* in the 3T3

620   NRU test method protocol recommended therein. The IC_{50} - LD_{50} regression obtained was

621   comparable to the RC and to the 11 chemical regression provided in the *Guidance Document*

622   (ICCVAM 2001b). The regression was log (LD_{50}) = 0.552 log IC_{50} + 0.503 (R^2=0.929)

623   while the RC regression was log (LD_{50}) = 0.435 log IC_{50} + 0.625. King and Jones (2003)

624   graphed the results to show that the regression fit within the acceptance interval ($\pm$ 0.5 log

625   around the regression line) of the RC.

626

627   King and Jones (2003) also showed that a 3T3 NRU test method altered for high throughput

628   testing by using a limited dose-response curve of three points yielded about the same IC_{50} as

629   an eight concentration dose-response. A regression used to compare the IC_{50} values using

630   the two different dose-response approaches yielded R^2 = 0.945.

631

632   9.3.3   A-Cute-Tox Project: Optimization and Pre-Validation of an *In Vitro* Test Strategy for

633   Predicting Human Acute Toxicity (Clemedson 2005)

634   The A-Cute-Tox Project is an Integrated Project under the EU 6th framework program that

635   started in January 2005 (termination date January 2010). The project was initiated in

636   response to the the REACH (Registration, Evaluation, Authorisation and Restriction of

637   Chemicals) Directive and the 7th amendment of the Cosmetics Directive call for the broad

638   replacement of animal experiments for finished products in 2003 and ingredients in 2009.

639   Dr. Cecilia Clemedson of Expertrådet Environmental Competence Ltd, Sweden, is the

640   scientific coordinator of the project.

641

642   The aim of the project is to develop a simple and robust *in vitro* testing strategy for

643   prediction of human acute systemic toxicity, which could replace the animal acute toxicity

644   tests used today for regulatory purposes. The objectives of A-Cute-Tox are:

645   • Compilation, critical evaluation, and generation of high quality *in vitro* and *in*

646   *vivo* data for comparative analysis.

647   • Identifying factors (kinetics, metabolism and organ specificity) that influence

648   the correlation between *in vitro* toxicity (concentration) and *in vivo* toxicity

649   (dosage), and to define an algorithm that accounts for this.

650   • Explore innovative tools and cellular systems to identify new end-points and

651   strategies to better anticipate animal and human toxicity.

652   • To design a simple, robust and reliable *in vitro* test strategy amenable for

653   robotic testing, associated with the prediction model for acute toxicity.

654

655   The project is an extension of the NICEATM/ECVAM study and the EDIT (Evaluation-

656   guided Development of *In-vitro* Test batteries) program, which is the continuation of the

657   MEIC (Multicentre Evaluation of *In Vitro* Cytotoxicity tests) study. The partnership is made

658   up of the EDIT Consortium, ECVAM, and 35 other European toxicity research group

659   partners. The project has been divided into the following workpackages that will be

660   implemented by various configurations of research partners:

661   • WP1: Generation of a "high quality" *in vivo* database (through literature

662   searches and historical data) and establishment of a depository list of reference

663   chemicals

664   • WP2: Generation of a "high quality" *in vitro* database (includes data from the

665   NICEATM/ECVAM study, EDIT studies, and MEIC studies)

666   • WP3: Iterative amendment of the testing strategy

667   • WP4: New end-points and new cell systems

668   • WP5: Alerts and correctors in toxicity screening (I): Role of ADE

669   • WP6: Alerts and correctors in toxicity screening (II): Role of metabolism

670   • WP7: Alerts and correctors in toxicity screening (III): Role of Target organ

671   toxicity (neuro-, nephro-, hepato-toxicity)

672   • WP8: Technical optimisation of the amended test strategy

673   • WP9: Pre-validation of the test strategy

674

675   A-Cute-Tox aims to extend the NICEATM/ECVAM and MEIC studies approach toward a

676   full replacement test strategy by improving the prediction of acute toxicity using *in vitro*

677   methods and validating the testing procedure.

678

679   **9.4   Summary**

680

681   • *In vitro* NRU cytotoxicity test methods using various cell types have been

682   evaluated for correlation with rodent lethality endpoints (e.g., rat/mouse iv, ip,

683   and oral toxicity). Peloux et al. (1992) and Fautrel et al. (1993) showed good

684   correlation (r=0.877 and 0.88, respectively) of *in vitro* cytotoxicity with rodent

685   ip/iv and iv data, respectively.

686   •   The 3T3 and NHK NRU test methods have been used for purposes other than

687   the prediction of starting doses for acute toxicity studies (e.g., ocular irritancy;

688   human lethal blood concentrations, *in vivo* phototoxicity).

689   •   The 3T3 NRU test method has been validated (through ECVAM) for the

690   identification of *in vivo* phototoxic potential.

691   •   No *in vitro* test methods have currently been validated for the prediction of

692   acute oral toxicity. Estimation of animal savings using *in vitro* cytotoxicity data

693   to estimate starting doses for the UDP did not use *in vitro* cytotoxicity data.

694   Instead, animal savings were estimated by assuming that the starting dose

695   equals the true LD_{50} (i.e., assumes cytotoxicity data can predict lethality

696   perfectly). Such theoretical predictions for animal savings for the UDP ranged

697   from 25-40% (ICCVAM 2001a) compared with the average animal savings of

698   6.6-13% predicted using computer simulation modeling of the UDP for the

699   chemicals tested in the NICEATM/ECVAM study. Halle et al. (1997) used the

700   *in vitro* cytotoxicity data in the RC to determine that animal savings of 32% can

701   be attained for the ATC method by using the LD_{50} predicted by the RC

702   regression as the starting dose. For the chemicals tested in the

703   NICEATM/ECVAM validation study, the average animal savings for the ATC,

704   determined by computer simulation modeling, was 8.0-14.8%.

705

706

36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51   ***[This Page Intentionally Left Blank]***
52
53

53 **10.0   ANIMAL WELFARE CONSIDERATIONS (REFINEMENT, REDUCTION,**
54 **   AND REPLACEMENT)**
55
56 As demonstrated in **Section 6**, *in vitro* NRU basal cytotoxicity test methods cannot be used
57 as replacement assays[1] for rodent acute oral toxicity test methods for hazard classification.
58 However, as described in this section, such test methods can be evaluated for their ability to
59 reduce[2] and refine[3] animal use in the UDP or ATC acute oral toxicity assays. A similar
60 analysis cannot be conducted for the FDP as this test method uses evident toxicity rather than
61 death as the endpoint of interest. The current UDP and ATC test guidelines recommend
62 using information on structurally-related substances and the results of any other toxicity tests
63 (EPA 2002b) to select a starting dose (OECD 2001a; EPA 2002a; OECD 2001d). However,
64 for the purposes of the reduction and refinement evaluation conducted in this section, it was
65 assumed that no information other than 3T3 and NHK NRU test data would be available
66 upon which to base the selection of a starting dose. To determine the extent of animal
67 reduction or refinement that would occur in the UDP and the ATC when using a starting dose
68 based on 3T3 or NHK NRU IC_{50} results rather than the default starting dose, computer
69 models were used to simulate the *in vivo* testing of the reference substances used in the
70 NICEATM/ECVAM validation study.
71
72 **Section 10.1** lists the regressions that were used with IC_{50} data from the 3T3 and NHK NRU
73 test methods to determine starting doses for the UDP and ATC test methods. **Sections 10.2.1**
74 and **10.3.1** summarize the animal testing procedures described in the current test guidelines
75 for the UDP and the ATC method, respectively. The procedures for using computer software
76 to simulate animal testing of the NICEATM/ECVAM reference substances are then detailed
77 in **Sections 10.2.2** and **10.3.2**. The computer simulations were used to determine the number
78 of animals used and the number of animals that died for each simulated test. The computer
79 simulation modelling was performed using five different dose-mortality (i.e., dose-response)

---

[1] **Replacement alternative:** A new or modified test method that replaces animals with nonanimal systems or one animal species with a phylogenetically lower one (e.g., a mammal with an invertebrate).
[2] **Reduction alternative:** A new or modified test method that reduces the number of animals required.
[3] **Refinement alternative:** A new or modified test method that refines procedures to lessen or eliminate pain or distress in animals or enhances animal well-being.

80 slopes since no information on dose-mortality slope was available for the substances tested.

81 To simplify the presentation of results, animal use figures provided in **Sections 10.2.3**,

82 **10.2.4**, **10.3.3**, and **10.3.4** include two of the dose-response slopes. The results for the other

83 three dose-response slopes are provided in **Appendices N** and **Q**. The number of animals

84 used is summarized to show the mean number of animals tested when the default starting

85 dose is used and the mean number of animals used when the NRU-determined starting dose

86 (i.e., from the 3T3 or NHK NRU IC_{50} values used in the indicated regressions) is used. The

87 difference in animal use between the default starting doses and the NRU-based starting doses

88 is referred to as the animal savings. Differences were tested for statistical significance (i.e., p

89 < 0.05) using a one-sided Wilcoxon signed ranked test based on the number of substances

90 evaluated. **Sections 10.2** and **10.3** summarize mean animal use by the total number of

91 substances tested and then by the number of substances in each GHS acute oral toxicity

92 category. **Sections 10.2.4** and **10.3.4** provide the mean number of animal deaths compared to

93 the mean number of animals used for each starting dose (i.e., default and NRU-based) to

94 determine whether the NRU-based starting doses result in the refinement of animal use (i.e.,

95 reduction in the number of animals that die).

96

97 **10.1   Use of 3T3 and NHK NRU Test Methods to Predict Starting Doses for Acute**

98 **Systemic Toxicity Assays**

99

100 The IC_{50} data from the 3T3 and NHK NRU test methods were used to predict starting doses

101 for acute oral systemic toxicity tests using the following linear regressions of IC_{50}-LD_{50}

102 values presented in **Section 6.2** (see **Table 6-2**):

103 •   the RC millimole regression [Note: The RC millimole regression was developed

104   from the Registry of Cytotoxicity, a database of rat and mouse oral LD_{50} values

105   from RTECS® and IC_{50} values from *in vitro* cytotoxicity assays using multiple

106   cell lines and cytotoxicity endpoints for 347 chemicals with known molecular

107   weights (Halle 1998).]

108 •   the RC rat-only weight regression

109 •   the RC rat-only weight regression excluding substances with specific

110   mechanisms of toxicity other than basal cytotoxicity

111   Data for the same reference substances were evaluated for each regression and simulated

112   acute systemic toxicity test method. Forty-six substances were evaluated for the 3T3 NRU

113   test method and 47 substances were evaluated for the NHK NRU test method. Of the 72

114   substances tested, epinephrine bitartrate, colchicine, and propylparaben were excluded

115   because they were removed from the calculation of the RC rat-only weight regression due to

116   the lack of rat oral reference LD_{50} data. The 21 substances with specific mechanisms of

117   toxicity in **Table 6-3** were excluded from all analyses to be consistent with those removed

118   from the RC rat-only weight regression excluding substances with specific mechanisms of

119   toxicity. These substances have known mechanisms of toxicity that are not expected to be

120   active in the 3T3 and NHK cell cultures. Carbon tetrachloride and methanol were excluded

121   from the 3T3 NRU evaluations because no laboratory attained sufficient toxicity in any test

122   for the calculation of an IC_{50}. Carbon tetrachloride was also excluded from the NHK NRU

123   evaluations because no laboratory attained sufficient toxicity in any test for the calculation of

124   an $IC_{50.}$

125

126   **10.2   Reduction and Refinement of Animal Use for the UDP**

127

128   10.2.1   <u>Procedure for *In Vivo* Testing Using the UDP</u>

129   This section describes the general dosing procedure for the UDP assay (OECD 2001a; EPA

130   2002a). Although doses, time between doses, and dose progression may be adjusted as

131   necessary, the procedures described reflect the default guidance. Guidance on the type of

132   animals to use, animal housing, clinical observations, etc., are outside the scope of the

133   current discussion and are provided in the test guidelines (see **Appendix M**).

134

135   *Main Test*

136   The UDP is based on a staircase design in which single animals are dosed in sequence at 48-

137   hour intervals. The outcome of the first animal determines the dose of the next animal. If the

138   first animal dies or is in a moribund state, the dose administered to the next animal is lowered

139   by dividing the original dose by one-half log (i.e., 3.2, which is the default dose progression).

140   If the first animal survives, the dose administered to the next animal is increased by one-half

141   log times the original dose. A dose progression of one-half log unit corresponds to a dose-

142   mortality (also referred to as "dose-response) slope of 2. The default dose progression can be

143   adjusted if the analyst has prior information upon which to estimate a slope.

144

145   The current test guidelines recommend using information on structurally-related substances

146   and the results of any other toxicity tests (EPA 2002b) for the test substance, including *in*

147   *vitro* cytotoxicity results, to approximate the LD_{50} and the slope of the dose-response curve

148   (OECD 2001a; EPA 2002a). The starting dose is one dose progression step below the

149   analyst's best estimate of the LD_{50}, since the UDP test method has a bias toward the starting

150   dose (i.e., LD_{50} estimate tends to move toward the starting dose). The default starting dose of

151   175 mg/kg is used if there is no information on which to base a starting dose. The entire

152   default dosing scheme generally uses a dose progression of 3.2, is 1.75, 5.5, 17.5, 55, 175,

153   550, 1750, and 5000 mg/kg (EPA 2002a) or 1.75, 5.5, 17.5, 55, 175, 550, and 2000 mg/kg

154   (OECD 2001a). Dosing single animals in sequence proceeds until the first of three

155   conditions, referred to as stopping rules, is met:

156   • three consecutive animals survive at the upper limit (2000 or 5000 mg/kg)

157   • five reversals occur in any six consecutive animals tested

158   • four or more animals have followed the first reversal and the specified

159   likelihood-ratios exceed the critical value. For a wide variety of LD_{50} values

160   and dose-mortality slopes, this is satisfied with four to six animals after the first

161   reversal. Three likelihood values are calculated: a likelihood at an LD_{50} point

162   estimate (called the rough estimate or dose-averaging estimate); a likelihood at a

163   value below the point estimate (the point estimate divided by 2.5); and a

164   likelihood at a value above the point estimate (the point estimate multiplied by

165   2.5). The ratios of the likelihoods are examined to determine whether they

166   exceed a critical value.

167

168   If none of these conditions is met, dosing stops after 15 animals have been used.

169

170   *Limit Test*

171   The UDP test method guidelines include a limit test using three to five animals dosed

172   sequentially at 2000 mg/kg or 5000 mg/kg (OECD 2001a; EPA 2002a). The EPA guideline

173   for testing at a limit dose of 5000 mg/kg calls for proceeding to the main test if the first

174   animal dosed at 5000 mg/kg dies (EPA 2002a). If the first animal lives, however, two more

175   animals are dosed at 5000 mg/kg. If both animals live, then testing is terminated with

176   LD_{50} > 5000 mg/kg. If one or both animals die, then two more animals are dosed in

177   sequence. As soon as three animals survive, the test is terminated with the conclusion that

178   LD_{50} > 5000 mg/kg. However, as soon as three animals die, the main test is conducted.

179   The OECD guideline for testing at a limit dose of 2000 mg/kg calls for proceeding to the

180   main test if the first animal dosed at 2000 mg/kg dies (OECD 2001a). If the animal lives,

181   however, four more animals are sequentially dosed. Whenever three animals die, the main

182   test is performed. If three or more animals survive, testing is terminated with the conclusion

183   that the LD_{50} > 2000 mg/kg.

184

185   10.2.2   Procedure for Computer Simulation Modeling of the UDP

186   Two thousand simulations of UDP testing were run for each substance, *in vitro* NRU test

187   method, and dose-mortality slope. Because the analysis assumed there was no information

188   upon which to estimate a dose-response slope, the simulation modeling used the default dose

189   progression factor of 3.2. The simulations used 5000 mg/kg as the upper limit dose since this

190   upper limit is commonly used in the United States. If the NRU-based starting dose was

191   4000 mg/mg or greater, then testing proceeded per the limit test rather than the main test. If,

192   during the dose progression, the next highest dose to be administered was within 4000 mg/kg

193   or greater, then the limit dose of 5000 mg/kg was administered. In the case where a dose one

194   step below the NRU-estimated LD_{50} was used as the starting dose, the other doses

195   administered corresponded to the default doses specified in the test method guidelines

196   (OECD 2001a; EPA 2002a). The simulation modeling procedures also used a lower limit of

197   1 mg/kg. Thus, if the dose progression fell below 1 mg/kg, then a dose of 1 mg/kg was

198   administered. To estimate animal use by the default method, a starting dose of 175 mg/kg

199   was used; the other doses administered after the default starting dose corresponded to the

200   default doses specified in the test method guidelines (OECD 2001a; EPA 2002a).

201

202   The simulation process was performed using SAS[®] version 8 (SAS 1999) and implements the

203   distributional assumptions underlying the dose-mortality relationship. The lowest dose at

204   which an animal dies in response to the administration of a toxic substance varies from

205   animal to animal. For an entire population of animals, mortality is assumed to have a log-

206   normal distribution with the mean equal to the log of the true LD_{50}. Sigma (σ), the

207   variability of the simulated population, is the inverse of the slope of the dose-mortality curve.

208   Due to a lack of information for the real dose-mortality curves, the simulations assumed

209   several different values of the slope, but no corresponding changes were made in the dose

210   progression. Dose-mortality slopes of 0.5, 0.8, 2, 4, and 8.3 were chosen since these were

211   used in the simulation modeling studies that evaluated the current version of the UDP

212   guidelines (ICCVAM 2001c).

213

214   To model the variability of the NRU IC_{50} values within and between laboratories, the values

215   were log-transformed to normalize the distribution of values for each substance. The mean

216   and variance of these log-transformed values were used to generate a log-normal distribution

217   from which to randomly select an IC_{50} value. The selected NRU IC_{50} value was used with

218   the regressions in two different ways to determine starting doses. One method used the LD_{50}

219   estimated from the IC_{50} and the regression as the starting dose while the other method used

220   the closest default dose lower than the estimated LD_{50} as the starting dose. The results from

221   the latter method are presented in **Section 10.2** since it is the method recommended by the

222   EPA and OECD test guidelines (EPA 2002a; OECD 2001a). Moreover, the UDP is only

223   usable for regulatory purposes if the starting dose is set below the expected LD_{50}. The results

224   obtained when the LD_{50} estimated by the IC_{50} and the regression was used as the starting

225   dose are presented in **Appendix Q.**

226

227   The simulation procedure used the following steps for each substance:

228   1. The LD_{50} value (in mg/kg) from **Table 4-2** was entered as the true LD_{50} value

229   and the choices of assumed slope were entered as the true slope for the dose-

230   mortality curve.

231   2. An IC_{50} value was selected from a distribution identified by the mean and

232   variance of the IC_{50} values computed from the data to reflect that different

233   laboratories produce different IC_{50} values in different situations (see **Table 5-3**

234   for mean IC_{50} values and standard deviations).

235   3. The IC$_{50}$ value from Step 2 was used in the regression model being evaluated to
236   compute a predicted LD$_{50}$ value to use as the starting dose.
237   4. The dosing simulation was run three times: once with the default starting dose
238   of 175 mg/kg, once at the next default dose below the LD$_{50}$ estimated by the
239   regression being evaluated, and once at a dose equal to that of the LD$_{50}$
240   estimated by the regression being evaluated.
241   5. For each simulated trial (each substance and starting dose), the dosing
242   simulation works similarly. In each trial, the animals are dosed sequentially;
243   therefore for each animal(i) there is a corresponding dose(i) that is administered
244   to the animal. For the first animal in each trial, it is the starting dose for that
245   trial. For each subsequent animal, the dose is dependent on the previous dose
246   and the previous animal's response as described in **Section 10.2.1**. For
247   animal(i), the probability of response is computed with the cumulative log-
248   normal distribution at the dose administered. That is,
249            $P(response) = P(x < \log[dose(i)])$ where $x \sim N(\mu,\sigma)$ and μ is the log of the
250   true LD$_{50}$ value and σ is the inverse of the assumed slope of the dose-mortality
251   curve. This probability is used to sample one observation from a binomial
252   distribution with this probability of success.
253   6. Dosing simulation is stopped once one of the stopping rules is satisfied.
254
255   Steps 2-6 were repeated 2000 times in order to compute an average animal use for each
256   method evaluated.
257

258   10.2.3   <u>Animal Savings for the UDP When Using 3T3 and NHK NRU-Based Starting</u>
259   <u>Doses</u>
260   10.2.3.1 *The Effect of Dose-Response Slope on Animal Use*
261   As described in **Section 10.2.2**, the simulation modeling of animal use for the UDP assumed
262   five different dose-mortality slopes to assess animal use under various conditions of
263   population variability. **Table 10-1** shows that the number of animals used for the UDP
264   decreases with increasing slope for both the default starting dose and the NRU-determined
265   starting dose based on the RC millimole regression. The NRU-determined starting dose was

266 the next default dose lower than the regression-estimated LD_{50}. For example, since the LD_{50}

267 predicted for cadmium chloride by the 3T3 NRU IC_{50} with the RC millimole regression was

268 16 mg/kg, the starting dose was 1.75 mg/kg (i.e., the next default dose below the predicted

269 LD_{50}). This approach is consistent with the UDP test method guidelines (OECD 2001a; EPA

270 2002a) as a means for reducing the number of animals that might experience pain and

271 suffering from treatment (i.e., as a test method refinement). The approach also overcomes

272 the nonconservative bias of the UDP, which tends to yield an LD_{50} close to the starting dose.

273

274 **Table 10-1   Change in Animal Use[1] with Dose-Response Slope for the UDP[2]**

Dose-Response Slope	With Default Starting Dose[1,3]	With NRU-Based Starting Dose[1,4]	Animals Saved[5]
3T3 NRU Test Method			
0.5	10.30 ± 0.13	9.43 ± 0.15	0.88* (8.5%)
0.8	10.34 ± 0.17	9.36 ± 0.18	0.98* (9.4%)
2.0	9.77 ± 0.21	8.79 ± 0.22	0.97* (10.0%)
4.0	8.96 ± 0.25	8.03 ± 0.27	0.93* (10.4%)
8.3	8.11 ± 0.26	7.20 ± 0.30	0.91* (11.2%)
NHK NRU Test Method			
0.5	10.31 ± 0.12	9.57 ± 0.17	0.74* (7.1%)
0.8	10.38 ± 0.16	9.47 ± 0.19	0.91* (8.8%)
2.0	9.75 ± 0.20	8.93 ± 0.23	0.82* (8.4%)
4.0	8.94 ± 0.24	8.14 ± 0.28	0.80* (9.0%)
8.3	8.12 ± 0.25	7.33 ± 0.30	0.79* (9.7%)

275 [1]Numbers are mean numbers of animals with standard errors for 2000 simulations for 46 substances
276 for the 3T3 NRU test method and 47 substances for the NHK NRU test method. Although the
277 simulations used whole animals, averaging the results produced fractional numbers of animals. The
278 slight differences in the number of animals used for the default starting dose at the same dose-
279 response slope reflect different simulation runs. Limit dose = 5000 mg/kg.
280 [2]OECD (2001a); EPA (2002a).
281 [3]Default starting dose = 175 mg/kg.
282 [4]Starting dose = next lower default dose to NRU-predicted LD_{50}, which was calculated using the
283 geometric mean of the laboratory geometric mean NRU IC_{50} values in the RC millimole regression:
284 $\log LD_{50}$ (mmol/kg) = 0.435 $\log IC_{50}$ (mM) + 0.625.
285 [5]Difference between mean animal use with default starting dose and mean animal use with NRU-
286 based starting dose. All differences denoted by * were statistically significant (i.e., $p < 0.05$) by a
287 one-sided Wilcoxon signed rank test. Percentage difference is shown in parentheses.
288
289

290 **Table 10-1** shows that, for each dose-response slope, the mean number of animals saved was

291 statistically significant (i.e., $p < 0.05$) when compared to mean animal use for the default

292   starting dose. When expressed as a percentage of the default animal use, animal savings also

293   generally increased with increasing slope.

294

295   To simplify the presentation of animal savings and comparison of the various regressions and

296   starting doses, the results of subsequent analyses presented in **Section 10.2.3** will be limited

297   to slopes of 2 and 8.3. The slope of 2 is the default slope used for the calculation of LD_{50} by

298   the UDP method (OECD 2001a; EPA 2002a). Animal savings results for the other dose-

299   mortality slopes are presented in **Appendices N1-N3.** Although using the next lower default

300   dose to the NRU-determined LD_{50} value overcomes the bias of the UDP toward the starting

301   dose (OECD 2001a, EPA 2002a) and is the appropriate approach for regulatory use, animal

302   savings results using the estimated LD_{50} as the starting dose were also calculated (see

303   **Appendix Q**).

304

305   10.2.3.2 *Mean Animal Use from UDP Simulations for Testing the NICEATM/ECVAM*

306   *Reference Substances – Comparison of Regressions and 3T3 and NHK NRU Test*

307   *Methods*

308   **Table 10-2** shows the mean animal use for simulated UDP of the testing the set of

309   NICEATM/ECVAM reference substances described in **Section 10.1**. Mean animal use is

310   shown for default starting dose and for starting doses that were one default dose lower than

311   the LD_{50} predicted from the *in vitro* NRU test methods and the regressions (shown in **Table**

312   **6-2**) evaluated in **Section 6.3** for prediction of GHS acute oral toxicity category. The

313   difference in animal use between the two starting doses is the mean animal savings produced

314   by using the starting dose based on the *in vitro* NRU test methods. All differences (i.e., mean

315   animal savings) were statistically significant (i.e., p < 0.05) by a one-sided Wilcoxon signed

316   rank test. Mean animal savings ranged from 0.79 to 1.16 (8.4 to 12.7%) animals depending

317   upon the NRU test method, regression, and dose-response slope. The lowest mean animal

318   savings were obtained for the RC millimole regression (0.82 [8.4%] to 0.97 [10.0%] animals

319   for the various test methods and dose-response slopes) and the highest mean animal savings

320   were obtained with the RC rat-only regression excluding substances with specific

321   mechanisms of toxicity other than basal cytotoxicity (1.00 [12.2%] to 1.16 [11.8%] animals).

322 **Table 10-2**   **Mean Animal Use[1] for the UDP[2] Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with**

323   **Various Regressions**

Assay/Regression	With Default Starting Dose[3]	With NRU-Based Starting Dose[4]	Animals Saved[5]	With Default Starting Dose[3]	With NRU-Based Starting Dose[5]	Animals Saved[5]	Accuracy[6]
3T3 NRU Test Method	**Dose-Response Slope = 2**			**Dose-Response Slope = 8.3**			
RC millimole[6]	9.77 ± 0.21	8.79 ± 0.22	0.97* (10.0%)	8.11 ± 0.26	7.20 ± 0.30	0.91* (11.2%)	26%
RC rat-only weight[7]	9.79 ± 0.21	8.66 ± 0.22	1.13* (11.6%)	8.14 ± 0.25	7.11 ± 0.29	1.03* (12.7%)	35%
RC rat-only weight excluding substances with specific mechanisms of toxicity[8]	9.80 ± 0.20	8.64 ± 0.23	1.16* (11.8%)	8.16 ± 0.25	7.08 ± 0.31	1.08* (13.3%)	46%
NHK NRU Test Method	**Dose-Response Slope = 2**			**Dose-Response Slope = 8.3**			
RC millimole[6]	9.75 ± 0.20	8.93 ± 0.23	0.82* (8.4%)	8.12 ± 0.25	7.33 ± 0.30	0.79* (9.7%)	28%
RC rat-only weight[7]	9.77 ± 0.20	8.83 ± 0.23	0.94* (9.6%)	8.13 ± 0.25	7.25 ± 0.30	0.88* (10.9%)	30%
RC rat-only weight excluding substances with specific mechanisms of toxicity[8]	9.78 ± 0.20	8.73 ± 0.24	1.05* (10.7%)	8.15 ± 0.25	7.15 ± 0.32	1.00* (12.2%)	38%

324 [1]Numbers are mean numbers of animals and standard errors for 2000 simulations for each of 46 substances for the 3T3 NRU test method and 47 substances for
325 the NHK NRU test method. Although the simulations used whole animals, averaging the results produced fractional numbers of animals. The slight differences
326 in the number of animals used for the default starting dose at the same dose-response slope reflect different simulation runs.
327 [2]OECD (2001a); EPA (2002a).
328 [3]Default starting dose = 175 mg/kg.
329 [4]Starting dose = one default dose lower than the NRU-predicted LD_{50} calculated using the geometric mean of the laboratory geometric mean NRU IC_{50} values in
330 the specified regression.
331 [5]Difference between mean animal use with default starting dose and mean animal use with NRU-based LD_{50}. Differences denoted by * were statistically
332 significant (i.e., $p < 0.05$) by a one-sided Wilcoxon signed rank test. Percentage difference is shown in parentheses.
333 [6]Proportion of substances for which the GHS acute oral toxicity category (UN 2005) predicted by the *in vitro* NRU test methods matched the *in vivo* category
334 (from **Tables 6-4** to **6-6**).
335 [7]$\log LD_{50}$ (mmol/kg) = 0.435 log IC_{50} (mM) + 0.625.
336 [8]$\log LD_{50}$ (mg/kg) = 0.372 log IC_{50} (µg/mL) + 2.024.
337 [9]$\log LD_{50}$ (mg/kg) = 0.357 log IC_{50} (µg/mL) + 2.194.
338
339

340   **Table 10-2** also shows that animal savings increased with the accuracy of the GHS acute oral

341   toxicity category predictions (see **Section 6.3**).

342

343   10.2.3.3 *Animal Savings for the UDP by Toxicity Category Using 3T3 and NHK NRU-Based*

344   *Starting Doses*

345   **Tables 10-3** through **10-5** show mean animal use and mean animal savings for the UDP for

346   the default starting dose and the NRU-determined starting dose with the test substances

347   grouped by GHS acute oral toxicity category (UN 2005). The data come from the same

348   analyses as the data provided in **Table 10-2**. NRU-determined starting doses were based on

349   the:

350   • RC millimole regression (**Table 10-3**).

351   • RC rat-only weight regression (**Table 10-4)**

352   • RC rat-only weight regression excluding substances with specific mechanisms

353   of toxicity other than basal cytotoxicity (**Table 10-5)**

354

355   Consistencies noted in the mean animal savings data provided in the tables included:

356   • For each *in vitro* NRU cytotoxicity test method and regression, animal savings

357   were statistically significant for substances in the $2000 < LD_{50} \leq 5000$ mg/kg

358   and $LD_{50} > 5000$ mg/kg toxicity categories.

359   • For substances with $LD_{50} \leq 5$ mg/kg, the NHK NRU test method with each

360   regression used slightly more animals than the default method (i.e., mean

361   animal savings were negative). The 3T3 NRU test method produced

362   nonsignificant animal savings of 0.31 (2.9%) to 0.95 (8.1%) animal for these

363   substances.

364   For substances with $50 < LD_{50} \leq 300$ mg/kg, all test methods and regressions produced little

365   to no animal savings.

366

367   *Animal Savings for the UDP by Toxicity Category Using 3T3 and NHK NRU-Based Starting*

368   *Doses with the RC Millimole Regression*

369   **Table 10-3** shows the animal savings by GHS toxicity category for the *in vitro* NRU

370   cytotoxicity test methods used with the RC millimole regression. Mean animal savings were

371   statistically significant (i.e., $p < 0.05$) by a one-tailed Wilcoxon signed rank test for the

372   following GHS toxicity categories, test methods, and dose-response slopes:

373   • $5 < LD_{50} \leq 50$ mg/kg for the NHK NRU at dose-response slope = 2 (0.86 [9.2%]

374   animals)

375   • $2000 < LD_{50} \leq 5000$ mg/kg for both NRU test methods and both dose-response

376   slopes (1.25 [13.7%] to 1.52 [14.1%] animals)

377   • $LD_{50} > 5000$ mg/kg for both NRU test methods and both dose-response slopes

378   (1.35 [14.2%] to 1.70 [25.4%] animals)

379

380   For the 3T3 NRU and NHK NRU test methods, mean animal savings were similar for most

381   toxicity categories at both dose-response slopes, with the mean savings for the 3T3 NRU

382   slightly higher than that for the NHK NRU. For the dose-response slope of 2, mean animal

383   savings for the 3T3 NRU test method (for the various toxicity categories) ranged from -0.09

384   (-1.0%) to 1.54 (16.1%) animals while mean animal savings for the NHK NRU test method

385   ranged from -0.25 (-2.2%) to 1.45 (13.5%) animals. For the dose-response slope of 8.3,

386   animal savings for the 3T3 NRU test method ranged from 0.004 (0.05%) to 1.70 (25.4%)

387   animals while mean animal savings for the NHK NRU test method ranged from

388   -0.11 (-1.5%) to 1.45 (21.8%) animals.

389

390   For both *in vitro* NRU cytotoxicity test methods, no mean animal savings (≤ 0.09 animal)

391   were observed for substances with $50 < LD_{50} \leq 300$ mg/kg. This category includes the

392   default starting dose of 175 mg/kg. Animal savings were not expected for this category since

393   savings were determined by comparing animal use with the NRU-based starting dose with

394   animal use at the default starting dose. For the 3T3 NRU, no animal savings (-0.9 to 0.004

395   animals) were also observed for substances with $5 < LD_{50} \leq 50$ mg/kg. For the NHK NRU

396   test method, animal use actually increased slightly compared to the default starting dose

397   (-0.25 to -0.09 animals) for substances with $LD_{50} \leq 5$ mg/kg. Animal savings for relatively

398   high toxicity substances were noted for those in the $LD_{50} \leq 5$ mg/kg category for the 3T3

399   NRU (0.78 [7.3%] to 0.95 [8.1%] animals) and in the $5 < LD_{50} \leq 50$ mg/kg category for the

400   NHK NRU (0.86 [9.2%] to 0.87 [10.5%] animals). Only the 0.86 (9.2%) animal savings for

401   the dose-response slope of 2 (NHK NRU) were statistically significant.

402 **Table 10-3**   **Animal Use[1] for the UDP[2] by GHS Toxicity Category[3] Using Starting Doses Based on the 3T3 and NHK NRU**

403   **Test Methods with the RC Millimole Regression[4]**

Toxicity Category[3]	Number of Reference Substances	Dose-Response Slope = 2			Dose-Response Slope = 8.3			Accuracy[8]
		With Default Starting Dose[5]	With NRU-Based Starting Dose[6]	Animals Saved[7]	With Default Starting Dose[5]	With NRU-Based Starting Dose[6]	Animals Saved[7]	
3T3 NRU Test Method								
$LD_{50} \le 5$ mg/kg	7	11.76 ± 0.16	10.8 ± 0.64	0.95 (8.1%)	10.65 ± 0.48	9.87 ±0.74	0.78 (7.3%)	0%
$5 < LD_{50} \le 50$ mg/kg	6	9.06 ± 0.18	9.15 ± 0.72	-0.09 (-1.0%)	8.04 ± 0.24	8.04 ± 0.78	0.004 (0.05%)	17%
$50 < LD_{50} \le 300$ mg/kg	6	7.70 ± 0.23	7.61 ± 0.18	0.09 (1.2%)	6.63 ± 0.35	6.59 ± 0.26	0.03 (0.5%)	67%
$300 < LD_{50} \le 2000$ mg/kg	6	8.76 ± 0.34	7.91 ± 0.06	0.84 (9.6%)	7.30 ± 0.35	6.69 ± 0.20	0.61 (8.3%)	100%
$2000 < LD_{50} \le 5000$ mg/kg	11	10.75 ± 0.08	9.23 ± 0.20	1.52* (14.1%)	9.16 ± 0.26	7.81 ± 0.34	1.36* (14.8%)	0%
$LD_{50} > 5000$ mg/kg	10	9.59 ± 0.27	8.05 ± 0.39	1.54* (16.1%)	6.69 ± 0.37	4.99 ± 0.45	1.70* (25.4%)	10%
NHK NRU Test Method								
$LD_{50} \le 5$ mg/kg	7	11.54 ± 0.25	11.79 ± 0.50	-0.25 (-2.2%)	10.63 ± 0.49	10.72 ± 0.54	-0.09 (-0.8%)	0
$5 < LD_{50} \le 50$ mg/kg	6	9.34 ± 0.24	8.48 ± 0. 24	0.86* (9.2%)	8.22 ± 0.31	7.35 ± 0.36	0.87 (10.5%)	50%
$50 < LD_{50} \le 300$ mg/kg	6	7.82 ± 0.22	7.88 ± 0.26	-0.06 (-0.7%)	6.92 ± 0.38	7.02 ± 0.43	-0.11 (-1.5%)	50%
$300 < LD_{50} \le 2000$ mg/kg	6	8.74 ± 0.34	7.93 ± 0.06	0.81 (9.3%)	7.31 ± 0.34	6.71 ± 0.23	0.60 (8.2%)	100%
$2000 < LD_{50} \le 5000$ mg/kg	11	10.73 ± 0.08	9.29 ± 0.20	1.45* (13.5%)	9.13 ± 0.25	7.88 ± 0.33	1.25* (13.7%)	9%
$LD_{50} > 5000$ mg/kg	11	9.52 ± 0.28	8.17 ± 0.41	1.35* (14.2%)	6.64 ± 0.35	5.19 ± 0.44	1.45* (21.8%)	0%

404 [1]Numbers are mean numbers of animals used and standard errors for 2000 simulations for each substance with a limit dose of 5000 mg/kg. Although the simulations used
405 whole animals, averaging the results produced fractional numbers of animals. Results are provided for 46 substances in the 3T3 NRU test method and 47 substances in
406 the NHK NRU test method categorized using the initial LD_{50} values from **Table 3-2**. The slight differences in the number of animals used for the default starting dose at
407 the same dose-response slope reflect different simulation runs.
408 [2]OECD (2001a); EPA (2002a).
409 [3]GHS-Globally Harmonized System of Classification and Labelling of Chemicals with LD_{50} in mg/kg (UN 2005).
410 [4]RC millimole regression is log LD_{50} (mmol/kg) = 0.435 log IC_{50} (mM) + 0.625.
411 [5]Default starting dose = 175 mg/kg.
412 [6]Starting dose was one default dose lower than the predicted LD_{50} calculated using the geometric mean of the laboratory geometric mean NRU IC_{50} values in the RC
413 millimole regression.
414 [7]Difference between mean animal use with default starting dose and mean animal use with NRU predicted LD_{50}. Differences marked by * are statistically significant (p <
415 0.05) by a one-sided Wilcoxon signed rank test. Percentage difference shown in parentheses
416 [8]Proportion of substances for which the GHS acute oral toxicity category (UN 2005) predicted by the *in vitro* NRU test methods matched the *in vivo* category (from
417 **Table 6-4**).

418　　**Table 10-3** also shows that mean animal savings did not correlate with the accuracy of the

419　　GHS acute oral toxicity category predictions. Substances in categories with the lowest

420　　accuracy produced the highest animal savings. Accuracy was the lowest (0 - 10%) for GHS

421　　acute oral toxicity category prediction for substances with $LD_{50} > 5000$ mg/kg, but animal

422　　savings (1.35 - 1.70) were the highest. Animal savings (0.60 - 0.84 animals) for substances

423　　with $300 \leq LD_{50} \leq 2000$ mg/kg, which had 100% accuracy for GHS acute oral toxicity

424　　category prediction, were similar to animal savings (0.78 - 0.95 animals) for substances in

425　　the $LD_{50} < 5$ mg/kg category (for the 3T3 NRU), which had 0% accuracy. Perhaps the

426　　difference between the predicted starting dose and the true LD_{50} vs. the difference between

427　　the default starting dose and the true LD_{50} has more influence on animal savings that the

428　　accuracy of the LD_{50} prediction.

429

430　　*Animal Savings for the UDP by Toxicity Category Using 3T3 and NHK NRU-Based Starting*

431　　*Doses with the RC Rat-Only Weight Regression*

432　　**Table 10-4** shows the mean animal savings by GHS toxicity category for the *in vitro* NRU

433　　cytotoxicity test methods used with the RC rat-only weight regression. A comparison of

434　　mean animal savings, category for category, with the RC millimole regression, indicates that,

435　　in most cases, animal savings were slightly higher for the RC rat-only weight regression. For

436　　the RC rat-only weight regression, the mean differences between animal use for the default

437　　starting dose and mean animal use with the NRU-determined starting dose were statistically

438　　significant (i.e., $p < 0.05$) by a one-sided Wilcoxon signed rank test for the following GHS

439　　toxicity categories, test methods, and dose-response slopes:

440　　　　• $300 < LD_{50} \leq 2000$ mg/kg for the NHK NRU at dose-response slope = 2 (0.86

441　　　　　　[9.8%] animals)

442　　　　• $2000 < LD_{50} \leq 5000$ mg/kg for both NRU test methods and both dose-response

443　　　　　　slopes (1.50 [16.4%] to 1.91 [17.7%] animals)

444　　　　• $LD_{50} > 5000$ mg/kg for both NRU test methods and both dose-response slopes

445　　　　　　(1.45 [15.2%] to 1.73 [25.9%] animals)

446 **Table 10-4   Animal Use[1] for the UDP[2] by GHS Toxicity Category[3] Using Starting Doses Based on the NRU Test Methods**

447 **with the RC Rat-Only Weight Regression[4]**

Toxicity Category[3]	Number of Reference Substances	Dose-Response Slope = 2			Dose-Response Slope = 8.3			Accuracy[8]
		With Default Starting Dose[5]	With NRU-Based Starting Dose	Animals Saved[7]	With Default Starting Dose[5]	With NRU-Based Starting Dose	Animals Saved[7]	
3T3 NRU Test Method								
$LD_{50} \leq 5$ mg/kg	4	11.75 ± 0.16	10.85 ± 0.61	0.89 (7.6%)	10.66 ± 0.48	9.93 ± 0.71	0.73 (6.8%)	0%
$> 5 < LD_{50} \leq 50$ mg/kg	7	9.14 ± 0.17	8.80 ± 0.54	0.34 (3.7%)	8.12 ± 0.27	7.76 ± 0.59	0.36 (4.5%)	17%
$> 50 < LD_{50} \leq 300$ mg/kg	5	7.75 ± 0.22	7.60 ± 0.10	0.15 (1.9%)	6.71 ± 0.32	6.66 ± 0.23	0.05 (0.8%)	67%
$> 300 < LD_{50} \leq 2000$ mg/kg	9	8.75 ± 0.33	7.89 ± 0.07	0.86* (9.8%)	7.29 ± 0.35	6.68 ± 0.21	0.61 (8.4%)	100%
$> 2000 < LD_{50} \leq 5000$ mg/kg	9	10.81 ± 0.08	8.90 ± 0.28	1.91* (17.7%)	9.18 ± 0.26	7.48 ± 0.42	1.70* (18.5%)	0%
> 5000 mg/kg	12	9.59 ± 0.27	7.96 ± 0.40	1.63* (17.0%)	6.69 ± 0.37	4.96 ± 0.45	1.73* (25.9%)	10%
NHK NRU Test Method								
$LD_{50} \leq 5$ mg/kg	4	11.58 ± 0.23	11.66 ± 0.44	-0.08 (-0.7%)	10.66 ± 0.48	10.59 ± 0.53	0.07 (0.6%)	0
$> 5 < LD_{50} \leq 50$ mg/kg	7	9.33 ± 0.26	8.39 ± 0.27	0.94 (10.1%)	8.20 ± 0.31	7.36 ± 0.38	0.84 (10.3%)	50%
$> 50 < LD_{50} \leq 300$ mg/kg	5	7.84 ± 0.21	7.93 ± 0.25	-0.09 (-1.1%)	6.94 ± 0.37	7.09 ± 0.41	-0.15 (-2.2%)	50%
$> 300 < LD_{50} \leq 2000$ mg/kg	9	8.74 ± 0.34	7.92 ± 0.06	0.82 (9.3%)	7.31 ± 0.34	6.71 ± 0.23	0.60 (8.2%)	100%
$> 2000 < LD_{50} \leq 5000$ mg/kg	9	10.77 ± 0.07	9.07 ± 0.24	1.70* (15.8%)	9.14 ± 0.25	7.64 ± 0.37	1.50* (16.4%)	9%
$LD_{50} > 5000$ mg/kg	13	9.52 ± 0.28	8.07 ± 0.40	1.45* (15.2%)	6.64 ± 0.35	5.09 ± 0.42	1.55* (23.3%)	0%

448 [1]Numbers are mean number of animals used and standard errors for 2000 simulations for each substance with a limit dose of 5000 mg/kg. Although the simulations used
449 whole animals, averaging the results produced fractional numbers of animals. Results are provided for 46 substances in the 3T3 NRU test method and 47 substances in
450 the NHK NRU test method categorized using the reference LD_{50} values from **Table 4-2**. The slight differences in the number of animals used for the default starting dose
451 at the same dose-response slope reflect different simulation runs.
452 [2]OECD (2001a); EPA (2002a).
453 [3]GHS-Globally Harmonized System of Classification and Labelling of Chemicals with LD_{50} in mg/kg (UN 2005).
454 [4]From **Table 6-2**; log LD_{50} (mg/kg) = 0.372 log IC_{50} (µg/mL) + 2.024
455 [5]Default starting dose = 175 mg/kg.
456 [6]Starting dose was one default dose lower than NRU-predicted LD_{50} calculated using the geometric mean of the laboratory geometric mean NRU IC_{50} values in the RC
457 rat-only regression.
458 [7]Difference between mean animal use with default starting dose and mean animal use with NRU predicted LD_{50}. Differences marked by * were statistically significant
459 (i.e., p < 0.05) by a one-sided Wilcoxon signed rank test. Percent difference is shown in parentheses.
460 [8]Proportion of substances for which the GHS acute oral toxicity category (UN 2005) predicted by the *in vitro* NRU test methods matched the *in vivo* category (from
461 **Table 6-5**).

462   For the dose-response slope of 2, mean animal savings (for the various toxicity categories)

463   for the 3T3 NRU test method ranged from 0.15 (1.9%) to 1.91 (17.7%) animals while mean

464   animal savings for the NHK NRU test method ranged from -0.09 (-1.1%) to 1.70 (15.8%)

465   animals. For the dose-response slope of 8.3, animal savings for the 3T3 NRU test method

466   ranged from 0.05 (0.8%) to 1.73 (25.9%) animals while animal savings for the NHK NRU

467   test method ranged from -0.15 (-2.2%) to 1.55 (23.3%) animals.

468

469   For both *in vitro* NRU cytotoxicity test methods, no mean animal savings (≤ 0.15 animal)

470   were observed for substances with $50 < LD_{50} \leq 300$ mg/kg. This category includes the

471   default starting dose of 175 mg/kg. Animal savings were not expected for this category since

472   savings were determined by comparing animal use with the NRU-based starting dose with

473   animal use at the default starting dose. For the NHK NRU, no animal savings (-0.08 to 0.07

474   animals) were also observed for substances with $LD_{50} \leq 5$ mg/kg. Animal savings for

475   relatively high toxicity substances were noted in the $LD_{50} \leq 5$ mg/kg category for the 3T3

476   NRU (0.73 [6.8%] to 0.89 [7.6%] animals) and in the $5 < LD_{50} \leq 50$ mg/kg category for the

477   NHK NRU (0.84 [10.3%] to 0.94 [10.1%] animals), but these savings were not statistically

478   significant.

479

480   **Table 10-4** also shows that mean animal savings did not correlate with the accuracy of the

481   GHS acute oral toxicity category predictions (see **Section 6.3**). The toxicity categories with

482   the highest animal savings had low accuracy. Substances in the $2000 < LD_{50} \leq 5000$ mg/kg

483   and $LD_{50} > 5000$ mg/kg categories had very low accuracy (0 - 10%) for GHS acute oral

484   toxicity category prediction, but the animal savings were higher than for the other categories

485   (1.45-1.91). Additionally, animal savings (0.61 - 0.86 animals) for substances with

486    $300 \leq LD_{50} \leq 2000$ mg/kg, which had 100% accuracy for GHS acute oral toxicity category

487   prediction, were similar to animal savings (0.73 - 0.89 animals) for substances in the $LD_{50} <$

488   5 mg/kg category (for the 3T3 NRU), which had 0% accuracy. Perhaps the difference

489   between the predicted starting dose and the true LD_{50} vs. the difference between the default

490   starting dose and the true LD_{50} has more influence on animal savings than the accuracy of the

491   LD_{50} prediction.

492   *Animal Savings for the UDP by Toxicity Category Using 3T3 and NHK NRU-Based Starting*

493   *Doses with the RC Rat-Only Weight Regression Excluding Substances with Specific*

494   *Mechanisms of Action*

495   **Table 10-5** shows the mean animal savings by GHS toxicity category for the *in vitro* NRU

496   cytotoxicity test methods used with the RC rat-only weight regression excluding substances

497   with specific mechanisms of toxicity other than basal cytotoxicity. For substances in the

498   categories for $LD_{50} > 2000$ mg/kg, mean animal savings for the RC rat-only weight

499   regression excluding substances with specific mechanisms of toxicity other than basal

500   cytotoxicity were slightly higher than those for the RC rat-only weight regression and those

501   for the RC millimole regression. Mean differences between animal use for the default

502   starting dose and mean animal use with the NRU-determined starting dose were statistically

503   significant (i.e., $p < 0.05$) by a one-sided Wilcoxon signed rank test for the following GHS

504   toxicity categories, test methods, and dose-response slopes:

505   • $5 < LD_{50} \le 50$ mg/kg for the NHK NRU at dose-response slope = 2 (0.98

506   [10.6%] animals)

507   • $300 < LD_{50} \le 2000$ mg/kg for both NRU test methods and at dose-response = 2

508   (1.00 [11.4%] animals for the 3T3 NRU and 0.90 [10.3%] animals for the NHK

509   NRU)

510   • $2000 < LD_{50} \le 5000$ mg/kg for both NRU test methods and both dose-response

511   slopes (1.75 [19.1%] to 2.22 [20.5%] animals)

512   • $LD_{50} > 5000$ mg/kg for both NRU test methods and both dose-response slopes

513   (1.77 [18.6%] to 2.01 [30.1%] animals)

514

515   Mean animal savings for the 3T3 NRU and NHK NRU test methods were similar for each

516   toxicity category and dose-response slope, with the 3T3 NRU test method producing slightly

517   higher mean animal savings in most cases. For the dose-response slope of 2, mean animal

518   savings across the various toxicity categories for the 3T3 NRU ranged from -0.02 (-0.2%) to

519   2.22 (20.5%) animals while mean animal savings for the NHK NRU ranged from -0.35

520   (-3.0%) to 1.98 (18.3%) animals.

521 **Table 10-5   Animal Use[1] for the UDP[2] By GHS Toxicity Category[3] Using Starting Doses Based on the 3T3 and NHK**

522 **NRU Test Methods with the RC Rat-Only Weight Regression Excluding Substances with Specific Mechanisms**

523 **of Toxicity[4]**

Toxicity Category[3]	Number of Reference Substances	Dose-Response Slope = 2			Dose-Response Slope = 8.3			Accuracy[8]
		With Default Starting Dose[5]	With NRU-Based Starting Dose	Animals Saved[7]	With Default Starting Dose[5]	With NRU-Based Starting Dose	Animals Saved[7]	
3T3 NRU Test Method								
$LD_{50} \le 5$ mg/kg	4	11.68 ± 0.17	11.26 ± 0.55	0.42 (3.6%)	10.62 ± 0.48	10.31 ± 0.67	0.31 (2.9%)	0%
> 5 < $LD_{50} \le 50$ mg/kg	7	9.05 ± 0.13	9.03 ± 0.55	0.02 (0.3%)	8.07 ± 0.25	7.92 ± 0.59	0.15 (1.9%)	14%
> 50 < $LD_{50} \le 300$ mg/kg	5	7.82 ± 0.18	7.84 ± 0.15	-0.02 (-0.2%)	6.93 ± 0.31	6.99 ± 0.29	-0.06 (-0.9%)	80%
> 300 < $LD_{50} \le 2000$ mg/kg	9	8.81 ± 0.35	7.81 ± 0.06	1.00* (11.4%)	7.31 ± 0.37	6.58 ± 0.18	0.73 (10.0%)	78%
> 2000 < $LD_{50} \le 5000$ mg/kg	9	10.84 ± 0.07	8.62 ± 0.23	2.22* (20.5%)	9.18 ± 0.26	7.19 ± 0.37	2.00* (21.8%)	67%
> 5000 mg/kg	12	9.59 ± 0.27	7.71 ± 0.40	1.88* (19.6)%	6.69 ± 0.37	4.68 ± 0.46	2.01* (30.1%)	25%
NHK NRU Test Method								
$LD_{50} \le 5$ mg/kg	4	11.55 ± 0.23	11.90 ± 0.32	-0.35(-3.0%)	10.66 ± 0.48	10.83 ± 0.45	-0.18 (-1.6%)	0
> 5 < $LD_{50} \le 50$ mg/kg	7	9.28 ± 0.25	8.30 ± 0.28	0.98* (10.6%)	8.19 ± 0.32	7.30 ± 0.36	0.89 (10.9%)	14%
> 50 < $LD_{50} \le 300$ mg/kg	5	7.87 ± 0.20	8.03 ± 0.24	-0.16 (-2.0%)	7.08 ± 0.34	7.26 ± 0.40	-0.19 (-2.6%)	60%
> 300 < $LD_{50} \le 2000$ mg/kg	9	8.76 ± 0.33	7.86 ± 0.06	0.90* (10.3%)	7.31 ± 0.34	6.61 ± 0.22	0.69 (9.5%)	89%
> 2000 < $LD_{50} \le 5000$ mg/kg	9	10.82 ± 0.07	8.84 ± 0.26	1.98* (18.3%)	9.15 ± 0.25	7.41 ± 0.39	1.75* (19.1%)	44%
$LD_{50} > 5000$ mg/kg	13	9.52 ± 0.28	7.75 ± 0.43	1.77* (18.6%)	6.64 ± 0.35	4.76 ± 0.44	1.88* (28.4%)	15%

524 [1]Numbers are mean number of animals used and standard errors for 2000 simulations for each substance with a limit dose of 5000 mg/kg. Although the
525 simulations used whole animals, averaging the results produced fractional numbers of animals. Results are provided for 46 substances in the 3T3 NRU test
526 method and 47 substances in the NHK NRU test method categorized using the reference LD_{50} values from **Table 4-2**. The slight differences in the number of
527 animals used for the default starting dose at the same dose-response slope reflect different simulation runs.
528 [2]OECD (2001a); EPA (2002a).
529 [3]GHS-Globally Harmonized System of Classification and Labelling of Chemicals with LD_{50} in mg/kg (UN 2005).
530 [4]From **Table 6-2**; log LD_{50} (mg/kg) = 0.357 log IC_{50} (µg/mL) + 2.194.
531 [5]Default starting dose = 175 mg/kg.
532 [6]Starting dose = One default dose lower than NRU-predicted LD_{50} calculated using the geometric mean of laboratory mean IC_{50} values in the RC rat-only weight
533 regression excluding substances with specific mechanisms of toxicity.

534   [7]Difference between mean animal use with default starting dose and mean animal use with NRU-based LD$_{50}$. Differences denoted by * were statistically
535   significant (i.e., p < 0.05) by a one-sided Wilcoxon signed rank test. Percent difference is shown in parentheses.
536   [8]Proportion of substances for which the GHS acute oral toxicity category (UN 2005) predicted by the *in vitro* NRU test methods matched the *in vivo* category
537   (from **Table 6-6**).
538

539   For the dose-response slope of 8.3, mean animal savings for the 3T3 NRU ranged from -0.06

540   (-0.9%) to 2.01 (30.1%) while mean animal savings for the NHK NRU ranged from -0.19

541   (-2.6%) to 1.88 (28.4%).

542

543   For both *in vitro* NRU cytotoxicity test methods, no mean animal savings were observed for

544   substances with $50 < LD_{50} \leq 300$ mg/kg. In fact, slightly more animals were used than when

545   using the default starting dose (i.e., animal savings were negative; -0.02 to -0.16 animal).

546   Since this category includes the default starting dose of 175 mg/kg, animal savings were not

547   expected for this category since savings were determined by comparing animal use with the

548   NRU-based starting dose with animal use at the default starting dose. For the NHK NRU test

549   method, more animals were also used for substances with $LD_{50} \leq 5$ mg/kg (i.e. animal

550   savings were -0.18 to -0.35 animals). The exceptions for having little to no animal savings

551   for the high toxicity substances was for the substances in the $5 < LD_{50} \leq 50$ mg/kg category

552   for the NHK NRU (0.89 [10.9%] to 0.98 [10.6%] animals), but only the 0.98 animals at dose-

553   response = 2 was statistically significant.

554

555   **Table 10-5** also shows that mean animal savings did not correlate with the accuracy of the

556   GHS acute oral toxicity category predictions (see **Section 6.3**). The toxicity categories with

557   the highest animal savings had low accuracy. Substances with $LD_{50} > 5000$ mg/kg had

558   relatively low accuracy (15 - 25%) for GHS acute oral toxicity category prediction, but the

559   animal savings were relatively high (1.88 - 2.01 animals). For the NHK NRU, substances in

560   the $5 < LD_{50} \leq 50$ mg/kg category had very low accuracy (14%) for GHS acute oral toxicity

561   category prediction, but the animal savings were statistically significant (0.98 animals at

562   dose-response = 2). Possibly the difference between the predicted starting dose and the true

563   LD_{50} vs. the difference between the default starting dose and the true LD_{50} has more

564   influence on animal savings than the accuracy of the LD_{50} prediction. The RC rat-only

565   weight regression excluding substances with specific mechanisms of toxicity improved

566   accuracy (compared with the RC millimole regression) and animal savings for the GHS

567   toxicity categories for substances in the $2000 < LD_{50} \leq 5000$ mg/kg and $LD_{50} > 5000$ mg/kg

568   categories. For substances in the $2000 < LD_{50} \leq 5000$ mg/kg category, accuracy increased

569   from 0 - 9% (both *in vitro* test methods and dose-response slopes) to 44 - 67% and animal

570   savings increased from 1.25 -1.52 animals to 1.75 - 2.22 animals. For substances with

571   LD_{50} > 5000 mg/kg, accuracy increased from 0 - 10% (both *in vitro* NRU test methods and

572   dose-response slopes) to 15 - 25% and animal savings increased from 1.35 - 1.70 animals to

573   1.77 - 2.01 animals. The RC rat-only weight regression excluding substances with specific

574   mechanisms of toxicity, however, also improved animal savings for substances in the

575    $300 < LD_{50} \leq 2000$ mg/kg toxicity category while which accuracy was decreased compared

576   with the RC millimole regression. Animal savings for substances in the

577    $300 < LD_{50} \leq 2000$ mg/kg toxicity category improved from 0.60 - 0.84 animals (for both *in*

578   *vitro* NRU test methods and dose-response slopes) to 0.69 - 1.00 animals while accuracy

579   decreased from 100% to 78 - 89%.

580

581   10.2.4   <u>Refinement of Animal Use for the UDP When Using 3T3 and NHK NRU-Based</u>

582   <u>Starting Doses</u>

583   A test method refines animal use when it lessens or eliminates pain or distress in animals or

584   enhances animal well-being (ICCVAM 2003). This section evaluates whether the use of 3T3

585   and NHK NRU-based starting doses refines animal use by reducing the number of animals

586   that die (i.e., experience pain and distress) during UDP testing compared to the number of

587   animals that die when using the default starting dose of 175 mg/kg. **Table 10-6** reports the

588   refinement results for the UDP simulation modeling using the 5000 mg/kg limit dose. For

589   every regression evaluated, the mean number of deaths when using the NRU-based starting

590   doses was slightly lower than the mean number of deaths when using the default starting

591   dose by approximately 0.1 to 0.2 deaths. The percentage of deaths, however, was slightly

592   higher for the NRU-based starting doses than for the default starting dose since the total

593   number of animals used was lower for the NRU-based starting doses. In general, fewer

594   animals were used and fewer animals died when using an NRU-based starting dose compared

595   with use of the default starting dose.

596

596 **Table 10-6**   **Animal Deaths[1] for the UDP[2] Using Starting Doses Based on the 3T3 and**
597   **NHK NRU Test Methods**

Assay/Regression	With Default Starting Dose[3]			With NRU-Based Starting Dose[4]		
	Used	Dead	% Deaths	Used	Dead	% Deaths
3T3 NRU Test Method	**Dose-Response Slope = 2**					
RC millimole[5]	9.77	4.16	42.6%	8.79	3.95	44.9%
RC rat-only weight[6]	9.79	4.18	42.6%	8.66	3.91	45.2%
RC rat-only weight excluding substances with specific mechanisms of toxicity[7]	9.80	4.18	42.7%	8.64	4.03	46.6%
	Dose-Response Slope = 8					
RC millimole[5]	8.11	3.43	42.3%	7.20	3.26	45.3%
RC rat-only weight[6]	8.14	3.44	42.3%	7.11	3.24	45.6%
RC rat-only weight excluding substances with specific mechanisms of toxicity[7]	8.16	3.45	42.3%	7.08	3.34	47.2%
NHK NRU Test Method	**Dose-Response Slope = 2**					
RC millimole[5]	9.75	4.10	42.0%	8.93	3.96	44.3%
RC rat-only weight[6]	9.77	4.11	42.0%	8.83	3.93	44.5%
RC rat-only weight excluding substances with specific mechanisms of toxicity[7]	9.78	4.12	42.1%	8.73	3.99	45.8%
	Dose-Response Slope = 8					
RC millimole[5]	8.12	3.38	41.7%	7.33	3.26	44.5%
RC rat-only weight[6]	8.14	3.39	41.7%	7.25	3.24	44.7%
RC rat-only weight excluding substances with specific mechanisms of action[7]	8.15	3.40	41.7%	7.15	3.29	46.1%

598 [1]Numbers are mean numbers of animals used for 2000 simulations for each substance. Although the
599 simulations used whole animals, averaging the results produced fractional numbers of animals. Upper limit
600 dose = 5000 mg/kg. Results are provided for 46 substances in the 3T3 NRU and 47 substances in the NHK
601 NRU test methods.
602 [2]OECD (2001a); EPA (2002a).
603 [3]Default starting dose = 175 mg/kg.
604 [4]Starting dose was one default dose lower than NRU-predicted LD_{50} calculated using the geometric mean of
605 laboratory mean IC_{50} values in the regression specified.
606 [5]log LD_{50} (mmol/kg) = 0.435 log IC_{50} (mM) + 0.625
607 [6]log LD_{50} (mg/kg) = 0.372 log IC_{50} (μg/mL) + 2.024
608 [7]log LD_{50} (mg/kg) = 0.357 log IC_{50} (μg/mL) + 2.194

609

609 **10.3   Reduction and Refinement of Animal Use for the ATC**

610

611 10.3.1   Procedure for *In Vivo* Testing Using the ATC

612 This section describes the general dosing procedure for the conduct of the ATC assay (OECD

613 2001d). The purpose of the ATC is to classify a test substance into the appropriate GHS

614 category for acute oral toxicity for classification and labeling. This is done by estimating the

615 range of the LD_{50} values for a test substance rather than calculating a point estimate of the

616 LD_{50}. The time between doses is determined by the onset, duration, and severity of toxic

617 signs. Guidance on the type of animals to use, animal housing, clinical observations, etc.,

618 which are outside the scope of the current discussion, are provided in the test guidelines (See

619 **Appendix M**).

620

621 *Main Test*

622 The ATC is based on the stepwise administration of test substances to three animals at a time

623 at one of a number of fixed doses: 5, 50, 300, and 2000 mg/kg (and 5000 mg/kg, if

624 necessary). The starting dose is selected so that at least some of the animals die at that dose.

625 If no information on which to base a starting dose is available, the default starting dose of

626 300 mg/kg is used. The next step, which may be to stop testing, test at the same dose, test at

627 the next higher dose, or test at the next lower dose, is determined by the starting dose and the

628 outcome of the three animals tested at the starting dose. For example, if the starting dose is

629 300 mg/kg and two to three animals die or are in a moribund state, the next step is to

630 administer 50 mg/kg to three more animals. However, if zero to one animal dies at 300

631 mg/kg, three more animals are tested at 300 mg/kg. Most substances required two to four

632 dose steps for substance classification. See **Appendix M** for the outcome-based testing

633 sequence for each starting dose.

634

635 *Limit Test*

636 For test substances that are likely to be nontoxic, the ATC test method guideline includes a

637 limit test in which six animals (three animals per step) are tested at the limit dose of

638 2000 mg/kg or 5000 mg/kg (OECD 2001d).

639

640   10.3.2   Procedure for Computer Simulation Modeling of the ATC

641   The simulation process for the ATC was performed using MATLAB® (The MathWorks, Inc.

642   1996-2004) computational software, which is functionally comparable to SAS® version 8.

643   Two thousand simulations of ATC testing were run for each substance, NRU test method,

644   and dose-mortality slope using an upper limit dose of 2000 mg/kg. The simulation process

645   implements the distributional assumptions underlying the dose-mortality response. The

646   lowest dose at which an animal dies in response to the administration of a toxic substance

647   varies from animal to animal. For an entire population of animals, mortality is assumed to

648   have a log-normal distribution with the mean equal to the log of the true LD_{50}. Sigma (σ),

649   the variability of the simulated population, is the inverse of the slope of the dose-mortality

650   curve. For any given dose, the probability that an animal will die is computed by the

651   cumulative log-normal distribution:

652

653    $$\text{Probability (death)} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\log dose} e^{\frac{-(t - \log trueLD_{50})^2}{2\sigma^2}} \, dt$$

654

655   Due to a lack of information for the real dose-mortality curves, the simulations assumed

656   several different values of the slope (i.e., the inverse of σ). Dose-mortality slopes of 0.5, 0.8,

657   2, 4, and 8.3 were chosen to be comparable to those chosen for simulation modeling of the

658   UDP (see **Section 10.2.2**).

659

660   To model the variability of the NRU IC_{50} values within and between laboratories, the values

661   were log-transformed to normalize the distribution of values for each substance. The mean

662   and variance of these log-transformed values were used to generate a log-normal distribution

663   from which to randomly select an IC_{50} value.

664

665   The simulation procedure used the following steps for each substance:

666   1.   The LD_{50} value (in mg/kg) from **Table 4-2** was entered as the true LD_{50} value

667   and the choices of assumed slope were entered as the true slope for the dose-

668   mortality curve.

669   2.   An IC_{50} value was selected from a distribution identified by the mean and

670   variance of the IC_{50} values computed from the data to reflect that different

671   laboratories produce different IC_{50} values in different situations (see **Table 5-3**

672   for mean IC_{50} values and standard deviations).

673   3.   The IC_{50} value from Step 2 was used in the regression model being evaluated to

674   compute a predicted LD_{50} value to use as the starting dose.

675   4.   The dosing simulation (of 2000 iterations) was run twice: once with the default

676   starting dose of 300 mg/kg and once with a starting dose equal to the next fixed

677   dose below the LD_{50} estimated by the regression being evaluated (i.e., the NRU-

678   based starting dose). If the NRU-based starting dose was greater than the 2000

679   mg/kg limit dose, then testing proceeded using the 2000 mg/kg limit test rather

680   than the main test.

681   5.   For every dose group of three animals, one observation was sampled from a

682   binomial distribution with the probability of death calculated by the probability

683   equation for a population of three. The sampled value, referred to as N1,

684   indicates the number of animals, 0, 1, 2, or 3, in the dosing group that die.

685   6.   If N1 ≤ 1, step 4 is repeated with the same dose. Now the sampled value from

686   the binomial distribution is referred to as N2.

687   7.   If N2 ≤ 1 and the dose is the highest dose tested, or the dose has already been

688   decreased, the toxicity category is assigned and testing is terminated. If the

689   dose is not the highest dose tested, or if the dose has not been decreased, the

690   dose is increased to the next fixed dose and step 4 is repeated.

691   8.   If N1 > 1 or N2 > 2, and the dose is the lowest dose tested, or if the dose has

692   already been increased, the toxicity category is assigned and testing is

693   terminated. If the dose is not the lowest dose tested, or if the dose has not

694   already been increased, the dose is decreased to the next fixed dose and step 4 is

695   repeated.

696

696   10.3.3   Animal Savings for the ATC When Using 3T3 and NHK NRU-Based Starting Doses

697   10.3.3.1 *The Effect of Dose-Response Slope on Animal Use*

698   As described in **Section 10.3.2**, the simulation modeling of animal use for the ATC used five

699   different dose-mortality slopes to assess animal use under various conditions of population

700   variability. **Table 10-7** shows how animal use for the simulated ATC changes with dose-

701   response slope and mean animal use for ATC simulations when using the default starting

702   dose of 300 mg/kg and when using a starting dose that was one fixed dose lower than that

703   predicted by the 3T3 and NHK NRU IC_{50} values with the RC millimole regression. The

704   mean number of animals used for the ATC decreased slightly with increasing slope for both

705   the default starting dose and the NRU-determined starting dose.

706

707   **Table 10-7   Change in Animal Use[1] with Dose-Response Slope for the ATC[2]**

Dose-Response Slope	Default Starting Dose[1,3]	NRU-Based Starting Dose[1,4]	Animals Saved[5]
3T3 NRU Test Method			
0.5	11.10 ± 0.07	10.11 ± 0.24	0.99* (8.9%)
0.8	10.98 ± 0.10	9.95 ± 0.27	1.03* (9.4%)
2.0	10.90 ± 0.16	9.76 ± 0.33	1.13* (10.4%)
4.0	10.84 ± 0.19	9.66 ± 0.35	1.17* (10.8%)
8.3	10.81 ± 0.21	9.64 ± 0.36	1.17* (10.8%)
NHK NRU Test Method			
0.5	11.10 ± 0.07	10.07 ± 0.22	1.03* (9.3%)
0.8	11.00 ± 0.09	9.90 ± 0.24	1.10* (10.0%)
2.0	10.93 ± 0.16	9.72 ± 0.30	1.21* (11.1%)
4.0	10.87 ± 0.19	9.61 ± 0.32	1.26* (11.6%)
8.3	10.84 ± 0.21	9.57 ± 0.34	1.27* (11.7%)

708   [1]Numbers are mean numbers of animals used and standard errors for 2000 simulations for 46
709   substances for the 3T3 NRU test method and 47 substances for the NHK NRU test method.
710   Although the simulations used whole animals, averaging the results produced fractional numbers of
711   animals. Limit dose = 2000 mg/kg.
712   [2]OECD (2001d).
713   [3]Default starting dose = 300 mg/kg.
714   [4]Next fixed dose lower than the predicted LD_{50} calculated using the geometric mean of laboratory
715   mean IC_{50} values in the RC millimole regression: log LD_{50} (mmol/kg) = 0.435 log IC_{50} (mM) + 0.625.
716   [5]Difference between mean animal use with default starting dose and mean animal use with NRU-based
717   starting dose. Differences that were statistically significant (i.e., $p < 0.05$) by a one-sided Wilcoxon
718   rank test are noted by *. Percent difference is shown in parentheses.

719

720   The mean numbers of animals saved, which was statistically significant (i.e., $p < 0.05$ by

721   one-sided Wilcoxon signed rank tests) when compared with mean animal use for the default

722　dose, generally increased with increasing slope. To simplify the presentation of animal

723　savings and comparison of the various regressions and starting doses, future results in

724　**Section 10.3.3** will be shown only for dose-response slopes of 2 and 8.3. Results for the

725　other dose-mortality slopes are presented in **Appendices N4-N6.**

726

727　10.3.3.2 *Mean Animal Use for ATC Simulations of Testing the NICEATM/ECVAM*

728　　　　　　*Reference Substances – Comparison of Regressions and 3T3 and NHK NRU Test*

729　　　　　　*Methods*

730　**Table 10-8** shows the mean animal use for testing the NICEATM/ECVAM reference

731　substances using the simulated ATC method when the starting dose was the default starting

732　dose and when the starting dose was one fixed dose lower than that determined by the LD_{50}

733　predicted from the 3T3 and NHK NRU test methods and the regressions (shown in **Table 6-**

734　**2**) evaluated in **Section 6.3** for prediction of GHS acute oral toxicity category. The mean

735　difference in animal use between the two starting doses is the mean animal savings. All

736　mean differences (i.e., mean animal savings) were statistically significant (i.e., $p < 0.05$ using

737　one-sided Wilcoxon signed rank tests). Mean animal savings ranged from 1.13 (10.4%) to

738　2.28 (21.1%) animals depending upon the test method, regression, and dose-response slope.

739　The lowest mean animal savings were obtained for the RC millimole regression (1.13

740　[10.4%] to 1.27 [11.7%] animals) and the highest mean animal savings were obtained with

741　the RC rat-only regression excluding substances with specific mechanisms of toxicity (1.68

742　[15.4%] to 2.28 [21.1%] animals).

743

744　10.3.3.3 *Animal Savings for the ATC by Toxicity Category Using 3T3 and NHK NRU-Based*

745　　　　　　*Starting Doses*

746　**Tables 10-9** through **10-11** show mean animal use and mean animal savings for the ATC

747　when used with the *in vitro* NRU cytotoxicity test methods, organized by GHS toxicity

748　category (UN 2005), and when based on the:

749　　　　　• RC millimole regression (**Table 10-9**)

750　　　　　• RC rat-only weight regression (**Table 10-10)**

751　　　　　• RC rat-only weight regression excluding substances with specific mechanisms

752　　　　　　　of toxicity (**Table 10-11)**

753 **Table 10-8   Animal Use[1] for the ATC[2] Using Starting Doses Based on NRU Test Methods with Various Regressions**

Assay/Regression	With Default Starting Dose[3]	With NRU-Based Starting Dose[4]	Animals Saved[5]	With Default Starting Dose[3]	With NRU-Based Starting Dose[5]	Animals Saved[5]	Accuracy[6]
3T3 NRU Test Method	Dose-Response Slope = 2			Dose-Response Slope = 8.3			
RC millimole[7]	10.90 ± 0.16	9.76 ± 0.33	1.13* (10.4%)	10.81 ± 0.21	9.64 ± 0.36	1.17* (10.8%)	26%
RC rat-only weight[8]	10.90 ± 0.16	9.21 ± 0.31	1.68* (15.5%)	10.81 ± 0.21	8.84 ± 0.36	1.97* (18.2%)	35%
RC rat-only weight excluding substances with specific mechanisms of toxicity[9]	10.90 ± 0.16	9.00 ± 0.29	1.90* (17.4%)	10.81 ± 0.21	8.53 ± 0.33	2.28* (21.1%)	46%
NHK NRU Test Method	Dose-Response Slope = 2			Dose-Response Slope = 8.3			
RC millimole[7]	10.93 ± 0.16	9.72 ± 0.30	1.21* (11.1%)	10.84 ± 0.21	9.57 ± 0.34	1.27* (11.7%)	28%
RC rat-only weight[8]	10.93 ± 0.16	9.45 ± 0.30	1.49* (13.6%)	10.84 ± 0.21	9.22 ± 0.34	1.62* (14.9%)	30%
RC rat-only weight excluding substances with specific mechanisms of toxicity[9]	10.93 ± 0.16	9.25 ± 0.26	1.68* (15.4%)	10.84 ± 0.21	8.91 ± 0.31	1.94* (17.9%)	38%

754 [1]Numbers are mean numbers of animals used and standard errors for 2000 ATC simulations each for 46 substances for the 3T3 NRU test method and 47
755 substances for the NHK NRU test method. Limit dose = 2000 mg/kg.
756 [2]OECD (2001d).
757 [3]Default starting dose = 300 mg/kg.
758 [4]Starting dose was one fixed dose lower than NRU-predicted LD_{50} calculated using the geometric mean of laboratory mean IC_{50} values in the regression
759 specified.
760 [5]Difference between mean animal use with default starting dose and mean animal use with NRU-based LD_{50}. Percentage difference is shown in parentheses.
761 Differences marked by * were statistically significant (i.e., $p < 0.05$) using a one-sided Wilcoxon signed rank test.
762 [6]Proportion of substances for which the GHS acute oral toxicity category (UN 2005) predicted by the *in vitro* NRU test methods matched the *in vivo* category
763 (from **Tables 6-4** to **6-6**).
764 [7]log LD_{50} (mmol/kg) = 0.435 log IC_{50} (mM) + 0.625.
765 [8]log LD_{50} (mg/kg) = 0.372 log IC_{50} (µg/mL) + 2.024.
766 [9]log LD_{50} (mg/kg) = 0.357 log IC_{50} (µg/mL) + 2.194.

767   The summarized data come from the same analyses as the data provided in **Table 10-8**.

768

769   Consistencies noted in the mean animal savings data provided in the tables included:

770   • For each test method and regression, the highest mean animal savings were

771   generally in the $LD_{50} \leq 5$ mg/kg and $LD_{50} > 5000$ mg/kg toxicity categories.

772   • For each test method and regression, the lowest mean animal savings were in

773   the $50 < LD_{50} \leq 300$ mg/kg and $300 < LD_{50} \leq 2000$ mg/kg toxicity categories.

774

775   *Animal Savings for the ATC by Toxicity Category Using 3T3 and NHK NRU-Based Starting*

776   *Doses with the RC Millimole Regression*

777   **Table 10-9** shows the mean animal savings for the ATC by GHS toxicity category for the *in*

778   *vitro* NRU test methods used with the RC millimole regression. Mean differences between

779   animal use for the default starting dose and animal use with the NRU-determined starting

780   dose were statistically significant (i.e., $p < 0.05$) by a one-sided Wilcoxon signed rank test for

781   the following GHS toxicity categories, test methods, and dose-response slopes:

782   • $LD_{50} \leq 5$ mg/kg for the 3T3 NRU at both dose-response slopes (2.75 [29.5%] to

783   2.80 [31.1%] animals)

784   • $2000 < LD_{50} \leq 5000$ mg/kg for the 3T3 NRU at dose-response slope = 8 (0.35

785   [2.9%] animals) and for the NHK NRU at dose-response slope = 2 (0.38 [3.4%]

786   animals)

787   • $LD_{50} > 5000$ mg/kg for the 3T3 NRU at both dose-response slopes (2.32

788   [29.6%] and 2.46 [20.5%] animals) and for the NHK NRU at dose-response

789   slope = 2 (2.34 [19.7%] animals)

790

791   For the dose-response slope of 2, mean animal savings for the 3T3 NRU test method ranged

792   from -0.24 (-2.5%) to 2.75 (29.5%) animals while animal savings for the NHK NRU test

793   method ranged from -0.02 (-0.2%) to 2.43 (19.9%) animals. For the dose-response slope of

794   8.3, mean animal savings for the 3T3 NRU test method ranged from -0.47 (-5.1%) to 2.80

795   (31.1%) animals while mean animal savings for the NHK NRU test method ranged from

796   -0.23 (-2.4%) to 2.79 (23.0%) animals.

797 **Table 10-9   Animal Savings[1] for the ATC[2] by GHS Toxicity Category[3] Using Starting Doses Based on the 3T3 and NHK**

798 **NRU Test Methods with the RC Millimole Regression[4]**

Toxicity Category[3]	Number of Reference Substances	Dose-Response Slope = 2			Dose-Response Slope = 8.3			Accuracy[8]
		With Default Starting Dose[5]	With NRU-Based Starting Dose[6]	Animals Saved[7]	With Default Starting Dose[5]	With NRU-Based Starting Dose[6]	Animals Saved[7]	
3T3 NRU Test Method								
$LD_{50} \leq 5$ mg/kg	7	9.35 ± 0.11	6.60 ± 0.87	2.75* (29.5%)	9.00 ± 0.001	6.20 ± 0.88	2.80* (31.1%)	0%
$5 < LD_{50} \leq 50$ mg/kg	6	12.22 ± 0.05	11.12 ± 0.94	1.10 (9.0%)	12.13 ± 0.08	10.71 ± 1.00	1.42 (11.7%)	17%
$50 < LD_{50} \leq 300$ mg/kg	6	10.70 ± 0.37	10.01 ± 0.08	0.69 (6.5%)	9.72 ± 0.48	9.39 ± 0.16	0.32 (3.3%)	67%
$300 < LD_{50} \leq 2000$ mg/kg	6	9.79 ± 0.08	10.04 ± 0.14	-0.24 (-2.5%)	9.20 ± 0.11	9.67 ± 0.27	-0.47 (-5.1%)	100%
$2000 < LD_{50} \leq 5000$ mg/kg	11	11.18 ± 0.08	11.02 ± 0.13	0.16 (1.4%)	11.90 ± 0.04	11.55 ± 0.20	0.35* (2.9%)	0%
$LD_{50} > 5000$ mg/kg	10	11.90 ± 0.03	9.58 ± 0.91	2.32* (19.5%)	12.00 ± 0.000	9.54 ± 0.97	2.46* (20.5%)	10%
NHK NRU Test Method								
$LD_{50} \leq 5$ mg/kg	7	9.37 ± 0.12	7.62 ± 1.12	1.76 (18.7%)	9.00 ± 0.002	7.25 ± 1.04	1.75 (19.5%)	0%
$5 < LD_{50} \leq 50$ mg/kg	6	12.2 ± 0.04	9.77 ± 0.34	2.43 (19.9%)	12.14 ± 0.09	9.35 ± 0.18	2.79 (23.0%)	50%
$50 < LD_{50} \leq 300$ mg/kg	6	10.75 ± 0.39	10.32 ± 0.36	0.43 (4.0%)	9.74 ± 0.49	9.97 ± 0.78	-0.23 (-2.4%)	50%
$300 < LD_{50} \leq 2000$ mg/kg	6	9.79 ± 0.08	9.81 ± 0.08	-0.02 (-0.2%)	9.21 ± 0.13	9.28 ± 0.13	-0.06 (-0.7%)	100%
$2000 < LD_{50} \leq 5000$ mg/kg	11	11.19 ± 0.09	10.81 ± 0.27	0.38* (3.4%)	11.90 ± 0.04	11.17 ± 0.73	0.73 (6.2%)	9%
$LD_{50} > 5000$ mg/kg	11	11.92 ± 0.02	9.58 ± 0.85	2.34* (19.7%)	12.00 ± 0.000	9.52 ± 0.90	2.48 (20.6%)	0%

799 [1]Numbers are mean number of animals used and standard errors for 2000 simulations for each substance with a limit dose of 2000 mg/kg. Results are provided
800 for 46 substances in the 3T3 NRU test method and 47 substances in the NHK NRU test method categorized using the initial LD_{50} values from **Table 3-2**.
801 Although the simulations used whole animals, averaging the results produced fractional numbers of animals.
802 [2]OECD (2001d).
803 [3]GHS-Globally Harmonized System of Classification and Labelling of Chemicals with LD_{50} in mg/kg (UN 2005).
804 [4]RC millimole regression is log LD_{50} (mmol/kg) = 0.435 log IC_{50} (mM) + 0.625.
805 [5]Default starting dose = 300 mg/kg.
806 [6]Starting dose was the next fixed dose lower than the predicted LD_{50} from using the NRU IC_{50} in the RC millimole regression.
807 [7]Difference between mean animal use with default starting dose and mean animal use with NRU-based starting dose. Statistically significant differences
808 (i.e., p < 0.05) by a one-sided Wilcoxon signed rank test are noted by *. Percentage difference is shown in parentheses.
809 [8]Proportion of substances for which the GHS acute oral toxicity category (UN 2005) predicted by the *in vitro* NRU test methods matched the *in vivo* category
810 (from **Table 6-4**).
811 .

812   For both dose-response slopes, the mean animal savings using the 3T3 NRU test method was

813   lower than the mean animal savings using the NHK NRU test method for substances in four

814   of the six toxicity categories: $5 < LD_{50} \leq 50$ mg/kg; $3000 < LD_{50} \leq 2000$;

815     $2000 < LD_{50} \leq 5000$ mg/kg; and $LD_{50} > 5000$ mg/kg. Mean animal savings using the 3T3

816   NRU test method was higher than the mean animal savings using the NHK NRU test method

817   for substances in the other two toxicity categories: $LD_{50} \leq 5$ mg/kg and

818    $50 < LD_{50} \leq 300$ mg/kg. For the 3T3 NRU test method, the highest mean animal savings

819   occurred for substances in the category for $LD_{50} \leq 5$ mg/kg (23.2 [19.5%] animals at dose-

820   response slope = 2 and 2.46 [20.5%] animals at dose-response slope = 8.3). For the NHK

821   NRU test method, the highest mean animal savings occurred for substances in the category

822   for $5 < LD_{50} \leq 50$ mg/kg (2.43 [19.9%] animals at dose-response slope = 2 and 2.79 [23.0%]

823   animals at dose-response slope = 8.3); however, the animal savings were not statistically

824   significant.

825

826   For both test methods, the smallest mean animal savings (≤ 0.69) were observed for

827   substances with LD_{50} values between 50 and 2000 mg/kg. Since the default starting dose

828   was 300 mg/kg, little change in mean animal use was expected for substances in the

829    $50 < LD_{50} \leq 300$ mg/kg and $300 < LD_{50} \leq 2000$ mg/kg categories. For both test methods and

830   dose-response slopes, mean animal savings for the substances in the $50 < LD_{50} \leq 300$ mg/kg

831   category were -0.23 to 0.69 animals. For both test methods and dose-response slopes, there

832   were no animal savings for substances in the $300 < LD_{50} \leq 2000$ mg/kg category. In fact,

833   slight more animals were used for the NRU-based starting doses than for the default starting

834   dose (-0.02 to -0.47 animals).

835

836   **Table 10-9** also shows that mean animal savings did not correlate with the accuracy of the

837   GHS acute oral toxicity category predictions (see **Section 6.3**). The toxicity categories with

838   the highest animal savings had low accuracy. The 3T3 NRU test method produced the

839   highest animal savings (2.75 - 2.80) for substances with $LD_{50} \leq 5$ mg/kg, which had 0%

840   accuracy for GHS acute oral toxicity category prediction. Substances in the

841    $300 < LD_{50} \leq 2000$ mg/kg category had 100% accuracy for GHS acute oral toxicity category

842   prediction, but had no animal savings (≤ 0.2 animals). Possibly the difference between the

843   predicted starting dose and the true LD_{50} vs. the difference between the default starting dose

844   and the true LD_{50} has more influence on animal savings than the accuracy of the LD_{50}

845   prediction.

846

847   *Animal Savings for the ATC by Toxicity Category Using 3T3 and NHK NRU-Based Starting*

848   *Doses with the RC Rat-Only Weight Regression*

849   **Table 10-10** shows the animal savings for the simulation ATC method by GHS toxicity

850   category for the *in vitro* NRU cytotoxicity test methods used with the RC rat-only weight

851   regression. Mean animal savings were statistically significant (i.e., p < 0.05) by a one-tailed

852   Wilcoxon signed rank test for the following GHS toxicity categories, test methods, and dose-

853   response slopes:

854   • $LD_{50} \leq 5$ mg/kg for both NRU test methods and dose-response slopes (2.03

855   [21.9%] to 2.57 [28.5%] animals)

856   • $2000 < LD_{50} \leq 5000$ mg/kg for the 3T3 NRU test method at both dose-response

857   slopes (1.39 [12.4%] to 2.56 [21.5%] animals)

858   • $LD_{50} > 5000$ mg/kg for both NRU test methods and dose-response slopes (2.92

859   [24.5%] to 3.5 [29.2%] animals)

860

861   For the 3T3 NRU and NHK NRU test methods, mean animal savings were similar for most

862   toxicity categories at both dose-response slopes, with the mean savings for the 3T3 NRU

863   slightly higher than that for the NHK NRU for most toxicity categories. For the dose-

864   response slope of 2, mean animal savings for the 3T3 NRU test method (for the various

865   toxicity categories) ranged from -0.32 (-3.3%) to 32.8 (27.5%) animals while mean animal

866   savings for the NHK NRU test method ranged from 0.03 (0.3%) to 2.92 (24.5%) animals.

867   For the dose-response slope of 8.3, animal savings for the 3T3 NRU test method ranged from

868   -0.63 (-6.8%) to 3.50 (29.2%) animals while mean animal savings for the NHK NRU test

869   method ranged from -0.23 (-2.4%) to 3.12 (26.0%) animals.

870

871   For both test methods, there were no mean animal savings (≤ 0.03 animals) for substances

872   with LD_{50} values between 300 and 2000 mg/kg. For both test methods and dose-response

873   slopes, mean animal savings for the substances in the $50 < LD_{50} \leq 300$ mg/kg category were

874 **Table 10-10   Animal Savings[1] for the ATC[2] by GHS Toxicity Category[3] Using Starting Doses Based on the 3T3 and NHK**

875   **NRU Test Methods with the RC Rat-Only Weight Regression[4]**

Toxicity Category[3]	Number of Reference Substances	Dose-Response Slope = 2			Dose-Response Slope = 8.3			Accuracy[8]
		With Default Starting Dose[5]	With NRU-Based Starting Dose[6]	Animals Saved[7]	With Default Starting Dose[5]	With NRU-Based Starting Dose[6]	Animals Saved[7]	
3T3 NRU Test Method								
$LD_{50} \leq 5$ mg/kg	4	9.35 ± 0.11	6.83 ± 0.84	2.52* (27.0%)	9.00 ± 0.001	6.43 ± 0.85	2.57* (28.5%)	0%
$> 5 < LD_{50} \leq 50$ mg/kg	7	12.22 ± 0.05	10.33 ± 0.52	1.88 (15.4%)	12.13 ± 0.08	9.94 ± 0.54	2.20 (18.1%)	14%
$> 50 < LD_{50} \leq 300$ mg/kg	5	10.70 ± 0.37	9.94 ± 0.10	0.76 (7.1%)	9.72 ± 0.48	9.23 ± 0.12	0.48 (5.0%)	80%
$> 300 < LD_{50} \leq 2000$ mg/kg	9	9.79 ± 0.08	10.11 ± 0.29	-0.32 (-3.3%)	9.20 ± 0.11	9.83 ± 0.55	-0.63 (-6.8%)	78%
$> 2000 < LD_{50} \leq 5000$ mg/kg	9	11.18 ± 0.08	9.79 ± 0.47	1.39* (12.4%)	11.9 ± 0.04	9.34 ± 0.82	2.56* (21.5%)	44%
> 5000 mg/kg	12	11.90 ± 0.03	8.62 ± 0.94	3.28* (27.5%)	12.00 ± 0.00	8.50 ± 0.99	3.50* (29.2%)	0%
NHK NRU Test Method								
$LD_{50} \leq 5$ mg/kg	4	9.37 ± 0.12	7.32 ± 0.88	2.05* (21.9%)	9.00 ± 0.002	6.97 ± 0.81	2.03* (22.6%)	0%
$> 5 < LD_{50} \leq 50$ mg/kg	7	12.20 ± 0.04	9.72 ± 0.30	2.48 (20.3%)	12.14 ± 0.08	9.35 ± 0.17	2.79 (23.0%)	14%
$> 50 < LD_{50} \leq 300$ mg/kg	5	10.75 ± 0.39	10.30 ± 0.34	0.45 (4.2%)	9.74 ± 0.49	9.97 ± 0.78	-0.23 (-2.4%)	60%
$> 300 < LD_{50} \leq 2000$ mg/kg	9	9.79 ± 0.08	9.76 ± 0.08	0.03 (0.3%)	9.21 ± 0.13	9.20 ± 0.11	0.02 (0.2%)	89%
$> 2000 < LD_{50} \leq 5000$ mg/kg	9	11.19 ± 0.09	10.45 ± 0.40	0.73 (6.6%)	11.90 ± 0.04	10.55 ± 0.69	1.35 (11.3%)	11%
$LD_{50} > 5000$ mg/kg	13	11.92 ± 0.02	9.00 ± 0.88	2.92* (24.5%)	12.00 ± 0.00	8.88 ± 0.93	3.12* (26.0%)	8%

876 [1]Numbers are mean number of animals used and standard errors for 2000 simulations for each substance with a limit dose of 5000 mg/kg. Although the
877 simulations used whole animals, averaging the results produced fractional numbers of animals. Results are provided for 46 substances in the 3T3 NRU test
878 method and 47 substances in the NHK NRU test method categorized using the reference LD_{50} values from **Table 4-2**.
879 [2]OECD (2001d).
880 [3]GHS-Globally Harmonized System of Classification and Labelling of Chemicals with LD_{50} in mg/kg (UN 2005).
881 [4]From **Table 6-2;** log LD_{50} (mg/kg) = 0.372 log IC_{50} (µg/mL) + 2.024
882 [5]Default starting dose = 300 mg/kg.
883 [6]Starting dose was one fixed dose lower than the NRU-predicted LD_{50} calculated using the NRU IC_{50} in the RC rat-only weight regression.
884 [7]Difference between mean animal use with default starting dose and mean animal use with NRU-based LD_{50}. Differences marked by * were statistically
885 significant (i.e., $p < 0.05$) by a one-sided Wilcoxon signed rank test. Percentage difference is shown in parentheses.
886 [8]Proportion of substances for which the GHS acute oral toxicity category (UN 2005) predicted by the *in vitro* NRU test methods matched the *in vivo* category
887 (from **Table 6-5**).
888

889   also relatively small (-0.23 to 0.76) animals. Since the default starting dose was 300 mg/kg,

890   little change in mean animal use was expected for substances in the $50 < LD_{50} \leq 300$ mg/kg

891   and $300 < LD_{50} \leq 2000$ mg/kg categories.

892

893   **Table 10-10** also shows that mean animal savings did not correlate with the accuracy of the

894   GHS acute oral toxicity category predictions (see **Section 6.3**). The toxicity categories with

895   the highest animal savings had low accuracy. For example, animal savings for substances in

896   the $LD_{50} > 5000$ mg/kg category were 2.92 - 3.50 animals (for both *in vitro* NRU test

897   methods and dose-response slopes) and accuracy was 0 - 8%. In addition, substances in

898   toxicity categories with the lowest animal savings had the highest accuracy for GHS acute

899   oral toxicity category prediction. Substances in the $300 < LD_{50} \leq 2000$ mg/kg category had

900   relatively high accuracy for GHS acute oral toxicity category prediction (i.e., 78% for the

901   3T3 NRU and 89% for the NHK NRU), but had the lowest animal savings (≤ 0.45 animals).

902   Possibly the difference between the predicted starting dose and the true LD_{50} vs. the

903   difference between the default starting dose and the true LD_{50} has more influence on animal

904   savings than the accuracy of the LD_{50} prediction.

905

906   *Animal Savings for the ATC by Toxicity Category Using 3T3 and NHK NRU-Based Starting*

907   *Doses with the RC Rat-Only Weight Regression Excluding Substances with Specific*

908   *Mechanisms of Toxicity*

909   **Table 10-11** shows the animal savings by GHS toxicity category for simulated ATC testing

910   using the *in vitro* NRU cytotoxicity test methods with the RC rat-only weight regression

911   excluding substances with specific mechanisms of toxicity. Mean animal savings were

912   statistically significant (i.e., $p < 0.05$) by a one-tailed Wilcoxon signed rank test for the

913   following GHS toxicity categories, test methods, and dose-response slopes:

914   • $LD_{50} \leq 5$ mg/kg for the 3T3 NRU test method at dose-response slope = 8.3 (2.16

915   [24.0%] animals) and for the NHK NRU test method at dose-response slope = 2

916   (1.27 [13.5%] animals)

917   • $2000 < LD_{50} \leq 5000$ mg/kg for both NRU test methods and both dose-response

918   slopes (1.23 [11.0%] to 3.07 [25.8%] animals)

919   • LD$_{50}$ > 5000 mg/kg for both NRU test methods and both dose-response slopes

920   (3.79 [31.8%] to 4.04 [33.7%] animals)

921

922   For the 3T3 NRU and NHK NRU test methods, mean animal savings were similar for most

923   toxicity categories at both dose-response slopes, with the mean savings for the 3T3 NRU

924   slightly higher than that for the NHK NRU. For the dose-response slope of 2, mean animal

925   savings for the 3T3 NRU test method (for the various toxicity categories) ranged from 0.02

926   (0.2%) to 4.08 (34.3%) animals while mean animal savings for the NHK NRU test method

927   ranged from 0.00 (0.0%) to 3.79 (31.8%) animals. For the dose-response slope of 8.3, animal

928   savings for the 3T3 NRU test method ranged from -0.03 (-0.4%) to 4.38 (36.5%) animals

929   while mean animal savings for the NHK NRU test method ranged from -0.06 (-0.6%) to 4.04

930   (33.7%) animals.

931

932   For both test methods, there were no mean animal savings ($\leq$ 0.02 animals) for substances

933   with LD$_{50}$ values between 300 and 2000 mg/kg. For both test methods and dose-response

934   slopes, mean animal savings for the substances in the 50 < LD$_{50}$ $\leq$ 300 mg/kg category were

935   also relatively small (-0.06 to 0.79) animals. Since the default starting dose was 300 mg/kg,

936   little change in mean animal use was expected for substances in the 50 < LD$_{50}$ $\leq$ 300 mg/kg

937   and 300 < LD$_{50}$ $\leq$ 2000 mg/kg categories.

938

939   **Table 10-11** also shows that mean animal savings did not correlate with the accuracy of the

940   GHS acute oral toxicity category predictions (see **Section 6.3**). The toxicity category with

941   the highest animal savings (LD$_{50}$ > 5000 mg/kg) had low accuracy (15 - 25%). Substances in

942   the 300 < LD$_{50}$ $\leq$ 2000 mg/kg category had very high accuracy, 78-89%, but no animal

943   savings. Perhaps the difference between the predicted starting dose and the true LD$_{50}$ vs. the

944   difference between the default starting dose and the true LD$_{50}$ has more influence on animal

945   savings than the accuracy of the LD$_{50}$ prediction.

946 **Table 10-11 Animal Savings[1] for the ATC[2] By GHS Toxicity Category[3] Using Starting Doses Based on the 3T3 and NHK**

947   **NRU Test Methods with the RC Rat-Only Weight Regression Excluding Substances with Specific Mechanisms**

948   **of Toxicity[4]**

Toxicity Category[3]	Number of Reference Substances	Dose-Response Slope = 2			Dose-Response Slope = 8.3			Accuracy[8]
		With Default Starting Dose[5]	With NRU-Based Starting Dose[6]	Animals Saved[7]	With Default Starting Dose[5]	With NRU-Based Starting Dose[6]	Animals Saved[7]	
3T3 NRU Test Method								
$LD_{50} \leq 5$ mg/kg	4	9.35 ± 0.11	7.23 ± 0.83	2.12 (22.6%)	9.00 ± 0.001	6.84 ± 0.86	2.16* (24.0%)	0%
$> 5 < LD_{50} \leq 50$ mg/kg	7	12.22 ± 0.05	10.52 ± 0.50	1.70 (13.9%)	12.13 ± 0.08	10.18 ± 0.54	1.96 (16.1%)	14%
$> 50 < LD_{50} \leq 300$ mg/kg	5	10.70 ± 0.37	9.92 ± 0.09	0.79 (7.3%)	9.72 ± 0.48	9.24 ± 0.13	0.48 (4.9%)	80%
$> 300 < LD_{50} \leq 2000$ mg/kg	9	9.79 ± 0.08	9.77 ± 0.07	0.02 (0.2%)	9.20 ± 0.11	9.24 ± 0.13	-0.03 (-0.4%)	78%
$> 2000 < LD_{50} \leq 5000$ mg/kg	9	11.18 ± 0.08	9.50 ± 0.47	1.67* (15.0%)	11.90 ± 0.04	8.83 ± 0.82	3.07* (25.8%)	67%
> 5000 mg/kg	12	11.90 ± 0.03	7.82 ± 0.77	4.08* (34.3%)	12.00 ± 0.00	7.62 ± 0.82	4.38* (36.5%)	25%
NHK NRU Test Method								
$LD_{50} \leq 5$ mg/kg	4	9.37 ± 0.12	8.11 ± 0.65	1.27* (13.5%)	9.00 ± 0.002	7.76 ± 0.58	1.24 (13.8%)	0%
$> 5 < LD_{50} \leq 50$ mg/kg	7	12.20 ± 0.04	9.87 ± 0.33	2.33 (19.1%)	12.14 ± 0.09	9.52 ± 0.27	2.62 (21.6%)	14%
$> 50 < LD_{50} \leq 300$ mg/kg	5	10.75 ± 0.39	10.19 ± 0.26	0.55 (5.2%)	9.74 ± 0.49	9.80 ± 0.61	-0.06 (-0.6%)	60%
$> 300 < LD_{50} \leq 2000$ mg/kg	9	9.79 ± 0.08	9.79 ± 0.08	0.00 (0.0%)	9.21 ± 0.13	9.21 ± 0.12	0.01 (0.1%)	89%
$> 2000 < LD_{50} \leq 5000$ mg/kg	9	11.19 ± 0.09	9.96 ± 0.45	1.23* (11.0%)	11.90 ± 0.04	9.62 ± 0.80	2.28* (19.2%)	44%
$LD_{50} > 5000$ mg/kg	13	11.92 ± 0.02	8.13 ± 0.76	3.79* (31.8%)	12.00 ± 0.000	7.96 ± 0.81	4.04* (33.7%)	15%

949 [1]Numbers are mean number of animals used and standard errors for 2000 simulations for each substance with a limit dose of 2000 mg/kg. Although the
950 simulations used whole animals, averaging the results produced fractional numbers of animals. Results are provided for 46 substances in the 3T3 NRU test
951 method and 47 substances in the NHK NRU test method categorized using the reference LD_{50} values from **Table 4-2**.
952 [2]OECD (2001d).
953 [3]GHS-Globally Harmonized System of Classification and Labelling of Chemicals with LD_{50} in mg/kg (UN 2005).
954 [4]From **Table 6-2**; log LD_{50} (mg/kg) = 0.357 log IC_{50} (µg/mL) + 2.194.
955 [5]Default starting dose = 300 mg/kg.
956 [6]Starting dose was one fixed dose lower than the NRU-predicted LD_{50} calculated using the NRU IC_{50} in the RC rat-only weight regression excluding substances
957 with specific mechanisms of toxicity.
958 [7]Difference between mean animal use with default starting dose and mean animal use with NRU-based LD_{50}. Statistically significant differences (i.e., $p < 0.05$)
959 by a one-sided Wilcoxon signed rank test are noted by *. Percentage difference is shown in parentheses.

960   [8]Proportion of substances for which the GHS acute oral toxicity category (UN 2005) predicted by the *in vitro* NRU test methods matched the *in vivo* category
961   (from **Table 6-5**).

962   The RC rat-only weight regression excluding substances with specific mechanisms of

963   toxicity improved accuracy (compared with the RC millimole regression) and animal savings

964   for the GHS toxicity categories for substances in the $2000 < LD_{50} \leq 5000$ mg/kg and $LD_{50} >$

965   5000 mg/kg categories. For the $2000 < LD_{50} \leq 5000$ mg/kg category, accuracy improved

966   from 0 - 9% (both *in vitro* NRU test methods) to 44 - 67% and animal savings improved from

967   0.16 - 0.73 animals to 1.23 - 3.07 animals. For substances with $LD_{50} > 5000$ mg/kg,

968   accuracy improved from 0 - 10% (both *in vitro* NRU test methods) to 15 - 25% and animal

969   savings improved from 2.32 - 2.48 animals to 3.79 - 4.38 animals. Although the RC rat-only

970   weight regression excluding substances with specific mechanisms of toxicity had no animal

971   savings for substances in the $300 < LD_{50} \leq 2000$ mg/kg toxicity category (≤ 0.02 animals), it

972   produced a small improvement over the RC millimole regression since as high as 0.47 more

973   animals were used (compared with the default starting dose).

974

975   10.3.4   Refinement of Animal Use for the ATC when using 3T3 and NHK NRU-Based

976   Starting Doses

977   A test method refines animal use when it lessens or eliminates pain or distress in animals or

978   enhances animal well-being (ICCVAM 2003). This section evaluates whether the use of 3T3

979   and NHK NRU-based starting doses refines animal use by reducing the number of animals

980   that die during ATC testing compared to the number of animals that die when using the

981   default starting dose of 300 mg/kg. **Table 10-12** reports the refinement results for the ATC

982   simulation modeling using the 2000 mg/kg limit dose. For every regression evaluated, the

983   mean number of deaths when using the 3T3 and NHK NRU-based starting doses was less

984   than the mean number of deaths when using the default starting dose by approximately 0.6 to

985   0.7 deaths. For the RC millimole regression and the RC rat-only weight regression, the

986   percentage of deaths (compared with the number of animals used) was also slightly lower for

987   the NRU-based starting dose compared with the default starting dose. For the RC rat-only

988   weight regression excluding substances with specific mechanisms of action, the percentage

989   of deaths (compared to the total number of animals used) when using the 3T3 and NHK

990   NRU-based starting doses was about the same as the percentage of deaths when using the

991   default starting dose. In general, fewer animals were used with the NRU-based starting dose

992   and fewer animals died.

993   **Table 10-12   Animal Deaths[1] for the ATC[2] Using Starting Doses Based on the 3T3 and**

994   **NHK NRU Test Methods**

Assay/ Regression	Default Starting Dose[3]			NRU-Based Starting Dose[4]		
	Used	Dead	% Deaths	Used	Dead	% Deaths
3T3 NRU	**Dose-Response Slope = 2**					
RC millimole[5]	10.90	3.55	32.6%	9.76	2.87	29.4%
RC rat-only[6]	10.90	3.55	32.6%	9.21	2.82	30.6%
RC rat-only excluding substances with specific mechanisms of toxicity[7]	10.90	3.55	32.6%	9.00	2.92	32.4%
	Dose-Response Slope = 8.3					
RC millimole[5]	10.81	3.03	28.0%	9.64	2.38	24.7%
RC rat-only[6]	10.81	3.03	28.0%	8.84	2.33	26.3%
RC rat-only excluding substances with specific mechanisms of toxicity[7]	10.81	3.03	28.0%	8.53	2.42	28.3%
NHK NRU	**Dose-Response Slope = 2**					
RC millimole[5]	10.93	3.47	31.8%	9.72	2.82	29.0%
RC rat-only[6]	10.93	3.47	31.8%	9.45	2.78	29.4%
RC rat-only excluding substances specific mechanisms of toxicity[7]	10.93	3.47	31.8%	9.25	2.91	31.5%
	Dose-Response Slope = 8.3					
RC millimole[5]	10.84	2.97	27.4%	9.57	2.34	24.4%
RC rat-only[6]	10.84	2.97	27.4%	9.22	2.30	24.9%
RC rat-only excluding substances with specific mechanisms of toxicity[7]	10.84	2.97	27.4%	8.91	2.43	27.3%

995   [1]Numbers are mean numbers of animals used for 2000 simulations for each substance (46 substances in the 3T3
996   NRU test method and 47 substances in the NHK NRU test method). Although the simulations used whole
997   animals, averaging the results produced fractional numbers of animals. Upper limit dose = 2000 mg/kg.
998   [2]OECD (2001d).
999   [3]Default starting dose = 300 mg/kg.
1000   [4]Starting dose was one fixed dose lower than the NRU-predicted LD_{50}.
1001   [5]log LD_{50} (mmol/kg) = 0.435 log IC_{50} (mM) + 0.625.
1002   [6]log LD_{50} (mg/kg) = 0.372 log IC_{50} (µg/mL) + 2.024.
1003   [7]log LD_{50} (mmol/kg) = 0.357 log IC_{50} (mM) + 2.194.
1004

1005   **10.4   Summary**

1006

1007   Computer simulation modeling of UDP testing using the default dose progression shows that,

1008   for the subset of NICEATM/ECVAM reference substances evaluated, the prediction of

1009   starting doses using the 3T3 and NHK NRU test methods with the RC millimole regression

1010   resulted in the use of statistically ($p < 0.05$) fewer animals for UDP testing by an average of

1011   0.79 - 0.97 animals (8.4 - 11.2%) depending upon the *in vitro* NRU cytotoxicity test method

1012   and the dose-response slope (of 2 or 8.3) used. Mean animal savings improved to 1.00 to

1013   1.16 animals (10.7 - 13.3%) for the RC rat-only weight regression excluding substances with

1014   specific mechanisms of toxicity.

1015

1016   When reference substances were grouped by GHS toxicity category, there were no mean

1017   animal savings for simulated UDP testing for substances with $50 < LD_{50} \leq 300$ mg/kg.

1018   Statistically significant animal savings were for observed for substances with $2000 < LD_{50} \leq$

1019   5000 mg/kg and $LD_{50} > 5000$ mg/kg for both NRU test methods. When using the RC

1020   millimole regression, animal savings for these categories ranged from 1.25 to 1.70 animals

1021   (13.5 to 25.4%). Use of the RC rat-only weight regression excluding substances with

1022   specific mechanisms of toxicity improved animal savings for substances in these toxicity

1023   categories to 1.75 to 2.22 animals (18.3 to 30.1%). Using the 3T3 and NHK NRU IC_{50}

1024   values to estimate starting doses for the simulated UDP also resulted approximately 0.1 to 0.2

1025   fewer mean deaths compared with the use of the default starting dose.

1026

1027   Computer simulation modeling of ATC testing with GHS cut points shows that, for the

1028   reference substances tested in this validation study, the prediction of starting doses using the

1029   3T3 and NHK NRU test methods with the RC millimole regression resulted in the use of

1030   statistically ($p < 0.05$) fewer animals for ATC testing by an average of 1.13 to 1.27 animals

1031   (10.4 - 11.7%) depending upon the *in vitro* NRU cytotoxicity test method and the dose-

1032   response slope (of 2 or 8.3) used. Animal savings improved to a mean of 1.68 to 2.28

1033   animals (15.4 - 21.1%) for the RC rat-only weight regression excluding substances with

1034   specific mechanisms of toxicity.

1035

1036   When test substances were grouped by GHS toxicity category, mean animal savings for ATC

1037   testing using the RC millimole regression were statistically significant for the 3T3 NRU at

1038   both dose-response slopes for substances with $LD_{50} \leq 5$ mg/kg (2.75 - 2.80 animals [29.5 -

1039   31.1%]) and for substances with $LD_{50} > 5000$ mg/kg (2.32 [19.5%] - 2.46 [20.5%] animals).

1040   Mean ATC animal savings with the RC millimole regression were statistically significant

1041   with the NHK NRU at dose-response = 2 for substances with $2000 < LD_{50} \leq 5000$ mg/kg

1042   (0.38 [3.4%] animals) and for substances with $LD_{50} > 5000$ mg/kg (2.34 animals [19.7%]).

1043   Using the RC rat-only weight regression excluding substances with specific mechanisms of

1044   toxicity, statistically significant animal savings were observed for both test methods and dose

1045   response slopes for substances with $2000 < LD_{50} \leq 5000$ mg/kg (1.23 [11.0%] - 3.07 [25.8%]

1046   animals) and substances with $LD_{50} > 5000$ mg/kg (3.79 [31.8%] - 4.38 [36.5%] animals).

1047   Animal savings were also statistically significant for substances with $LD_{50} \leq 5$ mg/kg using

1048   the 3T3 NRU at dose-response slope = 8.3 (2.16 [24.0%]) and using the NHK NRU at dose-

1049   response slope = 2 (1.27 [13.5%)]. Using the NRU IC_{50} values to estimate starting doses for

1050   the ATC refined animal use by producing approximately 0.6 to 0.7 fewer mean animal deaths

1051   than when the default starting dose of 300 mg/kg was used.

1052

1053   Spielmann et al. (1999) indicated that 76% (845/1115) of the industrial substances submitted

1054   to the Federal Institute for Health Protection of Consumers and Veterinary Medicine in

1055   Berlin, Germany, since 1982 had $LD_{50} > 2000$ mg/kg. Thus, the selection of starting doses

1056   using the *in vitro* NRU methods may save a considerable number of animals since animal

1057   savings are highest for the least toxic substances. However, the extent to which these

1058   substances represent the world of substances in commerce is not known.

1059

1059

1060

1061

1062

1063

1064

1065

1066

1067

1068

1069

1070   ***[This Page Intentionally Left Blank]***

1071

1072

19
20
21
22
23
24
25
26
27
28
29
30
31
32
33 ***[This Page Intentionally Left Blank]***
34

34 **11.0   PRACTICAL CONSIDERATIONS**

35

36 The 3T3 and NHK NRU test methods are proposed as adjuncts, rather than replacements for,

37 the *in vivo* acute oral toxicity assays. Data from these *in vitro* basal cytotoxicity test methods

38 are used with a prediction model to estimate the rodent oral LD_{50} of the test chemical. This

39 LD_{50} value is then used to determine the starting dose for subsequent *in vivo* acute oral

40 toxicity assays. This section discusses practical issues involved in applying these two *in*

41 *vitro* NRU cytotoxicity test methods to the prediction of starting doses for rodent acute

42 systemic toxicity assays. Practical issues to consider for implementation of these cell culture

43 test methods include the need for and availability of specialized equipment, training and

44 expertise requirements, cost considerations, and time expenditure. Good Cell Culture

45 Practice: ECVAM Good Cell Culture Practice Task Force Report 1 (Hartung et al. 2002)

46 encourages the establishment of practices and principles that will reduce uncertainty in the

47 development and application of *in vitro* test methods.

48

49 Good cell culture practices (in conjunction with good laboratory practices) are essential for

50 all *in vitro* cytotoxicity testing and should be employed to ensure that data produced from the

51 3T3 and NHK NRU test methods are reproducible, reliable, credible, and acceptable.

52

53 **11.1   Transferability of the 3T3 and NHK NRU Test Methods**

54

55 Transferability of a test method is defined as the ability of a test method or procedure to be

56 accurately and reliably performed in different, competent laboratories (ICCVAM 2003).

57 Accuracy and reliability of these test methods are discussed in **Sections 6** and **7**, respectively.

58

59 Protocols for the 3T3 and NHK NRU test methods, solubility testing, and prequalification of

60 keratinocyte growth medium have been optimized and are available on the

61 ICCVAM/NICEATM website (http://iccvam.niehs.nih.gov/methods/invitro.htm). The

62 protocols were designed with GLP-compliance in mind and can be easily implemented or

63 adapted by scientists with the appropriate technical experience.

64

65   While *in vitro* and *in vivo* methods require some similar skills (e.g., preparation of solutions

66   and test chemical doses, documentation), *in vitro* testing requires skills specific to cell culture

67   systems (e.g., aseptic techniques, microscopic evaluation of cell cultures, propagation of cells

68   in medium) but not to the maintenance, handling, or treatment of rodents.

69

70   11.1.1   Facilities and Major Fixed Equipment

71   The following lists of facility requirements, equipment and supplies, and training and

72   expertise are common to most *in vitro* mammalian cell culture laboratories. Required

73   equipment and supplies are also described in the NICEATM/ECVAM validation study 3T3

74   and NHK NRU test method protocols (**Appendices B and C**), the *Guidance Document*

75   (ICCVAM 2001b, **Appendix D**) and Hartung et al. 2002.

76

77   *Facility Requirements*

78   The testing facility should provide structures and infrastructures necessary for operating a

79   scientific laboratory (e.g., laboratory space, access to utilities, shipping/receiving department

80   [for appropriate receipt and handling of cell culture materials], etc.). Each facility should

81   provide:

82   •   personnel that are competent in performing *in vitro* cytotoxicity assays under

83   aseptic laboratory conditions

84   •   adequate facilities, equipment, and supplies

85   •   proper health and safety guidelines

86   •   satisfactory quality assurance procedures

87

88   Each facility should conform to all appropriate statutes (i.e., local, state, provincial, federal,

89   national, international) concerning safety guidelines (e.g., general workplace safety

90   guidelines, chemical handling and disposal guidelines, biohazard guidelines, etc.). Hartung

91   et al. 2002 provides recommended safety guidelines for working with potentially infectious

92   materials (e.g., HIV, hepatitis B, hepatitis C) and human materials (e.g., cells, tissues, fluids).

93

94   The facility management should establish scientific guidelines and procedures, train and

95   supervise professional and technical staff, and evaluate results and performance within their

96   discipline area relative to the testing requirements. Personnel should have mandatory

97   training in basic cell culture practice, in specific procedures for specialized culture

98   procedures, and in specific safety practices appropriate to the types of materials that may be

99   used in the laboratory (Hartung et al. 2002). The management should maintain records of the

100 qualifications, training and experience, and job descriptions for each professional and

101 technical individual involved in the testing.

102

103 *Cell Culture Laboratory*

104 The testing facility should have a designated cell culture laboratory to ensure that *in vitro*

105 cytotoxicity assays are performed under clean and proper aseptic conditions. The laboratory

106 should be located such that through traffic is minimal to reduce possible disturbances that

107 may compromise the cell culture assays. Room temperature of the laboratory should be

108 regulated, monitored, and documented. Access to the laboratory and test chemicals should

109 be restricted to appropriate personnel.

110

111 *Major Equipment*

112 Each testing facility should have at a minimum the following equipment:

113   • incubator (37ºC ± 1ºC, 90% ± 10% humidity, 5.0% ± 1% CO_2/air)

114   • laminar flow clean bench/cabinet (standard: "biological hazard")

115   • inverse phase contrast microscope

116   • 96-well plate spectrophotometric plate reader equipped with 540 nm ± 10 nm

117   filter (if testing in 96-well plates)

118   • autoclave

119   • refrigerator

120   • freezer (-70ºC)

121   • liquid nitrogen

122   • cryogenic freezer/storage unit

123   • computer

124

125 Equipment maintenance and calibration should be routinely performed and documented as

126 per GLP guidelines and testing facility procedures.

127

128   11.1.2   <u>Availability of Other Necessary Equipment and Supplies</u>

129   *General Equipment*

130   Each testing facility should have at a minimum the following equipment:

131   • centrifuge

132   • waterbath

133   • pipettors

134   • balance

135   • pH meter

136   • cell counting system

137   • water bath sonicator

138   • magnetic stirrer

139   • vortex mixer

140   • antistatic bar ionizer

141

142   Equipment maintenance and calibration should be routinely performed and documented as

143   per GLP guidelines and testing facility procedures. These types of equipment are available

144   from scientific and laboratory supply companies (e.g., Fisher Scientific, Thomas Scientific,

145   etc.).

146

147   *General Cell Culture Materials and Supplies*

148   The following supplies are needed for the NRU test methods:

149   • tissue culture plasticware

150   • glassware

151   • sterile filtration systems

152   • culture medium and supplements

153   • serum

154   • balanced salt solutions

155   • NRU assay chemicals

156

157   Cell culture supplies are generally available through the major scientific and laboratory

158   supply companies and through specialty companies (e.g., GIBCO, SIGMA-Aldrich,

159   CAMBREX/Biowhittaker, Becton Dickinson, etc.). Compositions of culture media,

160   supplements/additives, salt solutions, NRU assay chemicals and the volumes needed for the

161   test methods should be defined. All culture vessels needed to assure proper cell propagation

162   should be defined.

163

164   During this study, obtaining an adequate supply of NHK medium was problematic for FAL.

165   Communication between the UK distributor and the laboratory was uneven and the SMT

166   intervened on several occasions in an attempt to resolve the supply issue. This illustrates the

167   need for additional sources of keratinocyte cell culture medium. Periodically, it was also

168   difficult to obtain NHK medium and supplements that adequately supported keratinocyte

169   growth similarly in all the laboratories. Although the purchased medium met the

170   manufacturer's QA/QC standards, certain lots of the medium and supplements did not

171   support the growth of NHK cells to the extent needed to meet the growth characteristics

172   required by the test method protocol. This necessitated the need to incorporate an NHK

173   medium prequalification protocol into the study. Prequalification of medium is

174   recommended to avoid unnecessarily repeating studies.

175

176   *Cell Cultures*

177   3T3 Mouse Fibroblasts: BALB/c 3T3 cells, clone 31, can be obtained from

178   national/international cell culture repositories (e.g., CCL-163, American Type Culture

179   Collection [ATCC], Manassas, VA).

180

181   Normal Human Epidermal Keratinocytes (NHK): non-transformed keratinocyte cells from

182   cryopreserved primary or secondary cells can be obtained from national/international cell

183   culture repositories (e.g., CAMBREX Bio Science, 8830 Biggs Ford Road, Walkersville,

184   MD) or isolated from donated tissue (using proper collection, preparation, and propagation

185   techniques).

186

187   Obtaining adequate supplies of keratinocytes may be difficult since preparing a pool of cells

188   depends on the availability of tissue donors. Procurement of a commercially available stock

189   pool of cells and storing them indefinitely in a cryogenics freezer is recommended.

190

191   **11.2   3T3 and NHK NRU Test Method Training Considerations**

192

193   11.2.1   Required Training and Expertise

194   Hartung et al. 2002 recommends that scientists involved in *in vitro* testing should have

195   training in basic cell culture aspects such as: sterile technique, handling culture media,

196   feeding cultures, cell counting, subculture (trypsinization), detection and elimination of

197   contamination, growth parameters, growth curves, viability assays, storage and

198   freezing/thawing of cells. Additionally, training is encouraged for special culture procedures

199   such as: primary cell and tissue cultures, toxicity testing, viability assays, cloning,

200   transfection, expression cloning, cell transformation and immortalization, and virus

201   propagation and isolation. Laboratory personnel should be trained in the application of GLP

202   requirements (see **Section 8.1.1**).

203

204   *Training and Expertise*

205   *In vitro* NRU cytotoxicity test methods require personnel trained specifically in sterile

206   tissue/cell culture techniques and general laboratory procedures. Performance of the test

207   methods requires a relatively moderate degree of technical capability and a high degree of

208   skill in monitoring and maintaining appropriate cell growth conditions, troubleshooting

209   potential and real problems in culture systems, and interpreting and analyzing cytotoxicity

210   data. Each individual engaged in the conduct of or responsible for the supervision of a study

211   shall have education, training, and experience, or combination thereof, to enable that

212   individual to perform the assigned duties. The NRU test methods do not require that

213   personnel be trained to perform *in vivo* testing.

214

215   *Specific Training and Expertise Needed for the In Vitro NRU Cytotoxicity Test Methods*

216   Personnel involved in performing the *in vitro* NRU cytotoxicity test methods should be well

217   experienced in general cell culture techniques and should be able to:

218   • work with cryogenic freezing apparatus

219   • pipette solutions with large volume pipettors and multi-channel pipettors

220   • establish cells in culture vessels under aseptic conditions and monitor growth;

221   recognize normal and abnormal cell growth characteristics; document

222   observations of cell cultures throughout all aspects of the cultures

223   • perform the *in vitro* assays by following the protocols to: grow the cells, treat

224   the cells with test chemicals, perform the NRU assay, measure endpoints (i.e.,

225   optical density measurements), transfer data to electronic templates

226   • operate equipment necessary for maintaining cell culture laboratories (e.g.,

227   incubators, biohazard hoods, spectrophotometric microtiter plate readers)

228

229   *General Laboratory Expertise Needed for the In Vitro NRU Cytotoxicity Test Methods*

230   Personnel should also be able to perform and understand basic laboratory techniques and

231   laboratory management:

232   • prepare cell culture solutions (e.g., culture medium, NRU solutions); measure

233   pH; know proper storage conditions and maintain proper documentation

234   • prepare test chemicals for application to cell culture test plates; follow solubility

235   protocols to adequately prepare test chemicals in solution; recognize solubility

236   issues (e.g., insolubility nature of chemical, precipitation) and implement

237   mechanical procedures for solubilizing the test chemicals

238   • monitor and control laboratory room conditions (e.g., temperature, humidity,

239   lighting, traffic); maintain equipment at conditions essential to cell cultures

240   (e.g., temperature, humidity, gas flow, calibrations)

241

242   *Personnel Needed to Perform the In Vitro NRU Cytotoxicity Test Methods*

243   • <u>Study Director</u>: the single point of study control; has the overall responsibility

244   for the technical conduct of the testing (e.g., GLP adherence); determines test

245   acceptance, provides SOPs, interprets and analyzes the data, documents testing

246   aspects, and produces all written reports.

247   • <u>Quality Assurance Officer</u>: monitors the testing to assure conformance with

248   GLP requirements; must be independent of the Study Director.

249   •   Laboratory Technician(s): individuals trained in sterile tissue/cell culture

250   techniques and general laboratory procedures and capable of performing the *in*

251   *vitro* NRU cytotoxicity test methods in a GLP-manner.

252

253   11.2.2   Training Requirements to Demonstrate Proficiency

254   Laboratories set their own criteria for proficiency, but in general, personnel should be able to

255   understand the protocol, carry out the protocol with guidance from an experienced

256   supervisor/trainer, and then carry out the protocol with no supervision. An experienced

257   supervisor determines when a technician is adequately trained since there is no precise level

258   of training that can be measured. Once the technician demonstrates competence in executing

259   all the aspects of the *in vitro* NRU cytotoxicity test method(s), it is appropriate to initiate

260   routine assessments of observations among personnel using a benchmark control test

261   substance (SLS for these two NRU test methods) to ensure consistency.

262

263   The laboratories in this study were experienced in performing *in vitro* cytotoxicity assays but

264   were required to train and develop additional skills through Phases I and II (e.g., data

265   collection and transfer to Excel® and PRISM® templates). Inexperienced laboratory

266   personnel were trained by completion of "training" NRU assays using SLS. In the early

267   phases of the ICCVAM/ECVAM validation study, the laboratories continued training by the

268   testing of coded reference chemicals of various toxicities and performing solubility testing on

269   the chemicals. This training improved proficiency among the staff of the laboratories for the

270   final phase of the validation study.

271

272   *GLP-Compliance Proficiency Criteria*

273   ECBC and IIVS conducted this study in compliance with GLP Standards (see **Section 8.1.1**).

274   The appropriate QA unit (as per GLPs) reviewed the various aspects of the study and issued a

275   QA statement that identified whether the methods and the results described in the Final

276   Report accurately followed the test method protocol and reflected the raw data produced

277   during the study, respectively, and provided assurance that all testing was done under the

278   principles of GLP. FAL (non GLP-adherent) followed GLP standards referenced in **Section**

279   **8.1.1** as guidelines for conducting this study. FAL had no QA unit to judge their compliance

280   with GLP guidelines.

281

282   **11.3   Test Method Cost Considerations**

283

284   11.3.1   3T3 and NHK NRU Test Methods

285   *Laboratory Costs*

286   Supplies such as cell culture chemicals, the reagents used to measure NRU, and cell culture

287   plasticware are available from numerous suppliers and are not cost prohibitive. Major

288   instruments and equipment that *in vitro* cytotoxicity laboratories need to implement the *in*

289   *vitro* NRU cytotoxicity test methods are described in **Section 11.1.1**.

290

291   The 3T3 NRU test method is generally less expensive to use than the NHK NRU test

292   method. One vial of the immortalized 3T3 cells ($180) can be propagated indefinitely by

293   passaging cells and periodically cryopreserving pools (i.e., numerous vials of cells). NHK

294   cells require a fresh sample of primary cells for each test run ($380 per vial). Since primary

295   NHK cells are only passaged once after initiating into culture, there are no cells available to

296   cryopreserve a stock pool of cells. The D-MEM medium used for the 3T3 cells is less

297   expensive, more "generic", and more readily available than keratinocyte-specific medium.

298   (See **Table 11-1**)

299

300

300   **Table 11-1   Costs for Cell Culture Materials and Commercial Laboratory *In Vitro***
301   **Cytotoxicity Testing**

Item	Cost (approximate)	Number of Tests Possible	Other
3T3 Cells	$180/vial[1]	indefinite	One vial can produce an indefinite supply of cells by propagating the cells in culture and periodically freezing a pool of cells.
NHK Cells	$380/vial[1]	~5 (96-well plates)	Since cells are passaged only once beyond cryopreservation, new ampules should be thawed frequently to maintain continuous testing.
Dulbeccos' Minimum Essential Medium (D-MEM) with supplements	$20/500mL[1]	~15 (96-well plates)	Establish cells in culture (~20 mL/vial of cells; 60 mL/3 vials), seed cells in 96-well plates (12 mL/plate; 180 mL/15 plates); prepare stock solution and eight concentration dilutions (~20 mL/chemical; 300 mL/15 plates).
NHK Medium with supplements	$80/500 mL[1]	~15 (96-well plates)	Same as DMEM (above)
Commercial Laboratory Testing (MB Research Laboratories)	$1050/$1950 (USP/ISO) per 3 test materials[2]	1 test/material	*in vitro* NRU cytotoxicity test (24-hour test period)
Commercial Laboratory Testing (Institute for In Vitro Sciences))	$1120 (GLP) per test material (minimum of 5 materials)[2]	1 range finder, 2 definitive tests per test material	*in vitro* NRU cytotoxicity test (48-hour test period)
Commercial Laboratory Testing (Institute for In Vitro Sciences))	$1850 (GLP) per single test material[2]	1 range finder, 2 definitive tests per test material	*in vitro* NRU cytotoxicity test (48-hour test period)

302   [1]catalogue price
303   [2]personal communication
304

305   *Commercial Testing Laboratories*

306   A representative of MB Research Laboratories (Spinnerstown, PA,

307   http://www.mbresearch.com/) provided a quote (personal communication 2005) for an *in*

308   *vitro* NRU cytotoxicity test (24-hour [and not a 48-hour] test period) of $1050/$1950

309   (USP/ISO) per set of three test chemicals. The lead laboratory for the NICEATM/ECVAM

310   study, IIVS (Gaithersburg, MD, http://www.iivs.org/) provides commercial laboratory GLP-

311   compliant testing using this study's protocols (48-hour test period) at a cost of $1120 - $1850

312   per chemical/sample (personal communication with Hans Raabe [IIVS] 2005).

313

314   11.3.2   *In Vivo* Rodent Acute Oral Toxicity Testing

315   **Table 11-2** provides commercial prices for acute oral systemic toxicity testing.

316

317   MB Research Laboratories performs the UDP test at a cost of $750 for three rats and charges

318    $250 for each additional rat needed. In the best-case scenario, the UDP test needs only three

319   rats ($750). In the worst-case scenario, this test would need an additional 12 rats (15

320   maximum for the test); the total cost of the test would be $3750. In this costing strategy,

321    $250 is saved from the total cost of the UDP for each rat saved by using the 3T3 or NHK

322   NRU test method to predict the starting dose. Considering that adding the *in vitro* NRU

323   cytotoxicity test costs from $350 to $1850 per chemical, the NRU test does not provide cost

324   savings if fewer than two to six animals are saved.

325

326   The President of Product Safety Laboratories (Dayton, NJ,

327   http://www.productsafetylabs.com/), Gary Wnorowski, provided a cost quote of $2700 for

328   determination of an LD_{50} value using the UDP test; the cost is independent of the number of

329   rats that are needed. Each testing dose is administered ~24-48 hours after the previous dose

330   and each animal test generally does not exceed four days. Time involved in providing the

331   LD_{50} value is approximately three months (initiation of the test to provision of the final

332   report). Knowing the estimated LD_{50} value does not affect the cost of the *in vivo* test in this

333   case but could reduce the number of animals needed for the test.

334

335   Bio Research Laboratories (BRL) performs the Acute Oral Rat Toxicity Test bioassay to

336   determine the relative acute toxicity of an unknown substance. The method determines

337   lethality and signs of acute toxicity from a waste sample administered in a single dose by

338   gavage to a limited number of rats. The bioassay determines if the test sample exhibits a

339   median lethal dose (LD_{50}) either greater than or less than a regulatory threshold

340   corresponding to a hazardous waste designation (i.e., 5000, 500, 50 mg/kg). A minimum of

341   ten rats is used at the tested dosage for the pertinent regulatory threshold value that is

342   relevant to the test sponsor. Knowledge of the estimated LD_{50} does not reduce animal use or

343   test costs if a single predetermined dose is tested.

344

345

345 **Table 11-2   Commercial Prices for Conducting *In Vivo* Acute Toxicity Testing**

Test	GLP-Compliant	Non GLP-Compliant	Company
Acute Oral Toxicity UDP: Limit Test - 2000 mg/kg	$1200	$1000	Product Safety Laboratories (PSL)
Acute Oral Toxicity UDP: Limit Test - 5000 mg/kg	$800	$650	PSL
Acute Oral Toxicity UDP: LD$_{50}$	$2700	$2200	PSL[a]
Acute Oral Rat Toxicity: single dose[b]	$950	NA	Bio Research Laboratories (BRL)
Acute Oral Rat Toxicity: two doses[b]	$1500	NA	BRL
Acute Oral Rat Toxicity: LD$_{50}$	$3000	NA	BRL
Acute Oral Toxicity – UDP	$730 for the first 3 animals; $250 each additional animal	NA	MB Research Laboratories[a]

346 [a]provided to NICEATM through personal communication
347 [b]Washington State Biological Testing Methods #80-12 For the Designation of Dangerous Waste; Part B: Acute
348 Oral Rat Toxicity Test [http://www.ecy.wa.gov/pubs/80012.pdf] The method is an adaptation of the EPA
349 Health Affects Test Guidelines OPPTS 870.110 Acute Oral Toxicity and American Society for Testing and
350 Materials (ASTM) methods E 1163-90 (Standard test method for estimating acute oral toxicity in rats) and E
351 1372-90 (Standard test method for conducting a 90-day oral toxicity study in rats).
352

353 **11.4   Time Considerations for the 3T3 and NHK NRU Test Methods**

354

355 *The 3T3 NRU Test Method*

356 Approximately one week is needed to thaw cryopreserved 3T3 cells, propagate the cells in

357 flasks, and passage/subculture the cells at least two times before subculturing to the 96-well

358 test plate. After subculture into 96-well plates, the cells are incubated another 24 hours to

359 reach the proper percentage of confluency, and then exposed to test chemical for 48 hours.

360 The entire 3T3 NRU assay process takes approximately 10 days. However, once the cells are

361 established in culture, they can be passaged for approximately two months before starting the

362 initial propagation from frozen stock. Multiple chemicals can be tested at the same time, and

363 different tests can overlap each other; thus, many chemicals can be tested in a relatively short

364 time.

365

366 *The NHK NRU Test Method*

367 Approximately one week is needed to thaw cryopreserved NHK cells, propagate the cells in

368 flasks, and passage/subculture the cells (<u>once</u>) directly to the 96-well test plate. After

369 subculture into 96-well plates, the cells are incubated another 48-72 hours to reach the proper

370   percentage of confluence and then exposed to test chemical for 48 hours. The entire NHK

371   NRU assay process (range finder or definitive test) requires approximately 11-12 days. Cells

372   can be seeded at different densities in the culture flasks so that passaging the cultures can

373   take place on different days. Once the cells are established in culture, they are passaged once

374   to the 96-well test plates. Multiple chemicals can be tested at the same time, and different

375   tests can overlap each other; thus, many chemicals can be tested in a relatively short time.

376

377   *In Vivo Testing*

378   According to guidelines for acute oral toxicity testing for the main test and limit dose test,

379   single animals or groups of animals are dosed in sequence, usually at 2-4 day intervals, and

380   observations are generally made for up to 14 days (for animals that are not moribund) (EPA

381   2002a; OECD 2001a; OECD 2001b, OECD 2001c). The addition of NRU testing to estimate

382   a starting dose prior to the implementation of the UDP main test or limit dose test will take

383   10-12 days, but could save up to 14 days of observation for every animal saved.

384

385   **11.5   Summary**

386   •   All equipment and supplies are readily available. Direct communication with

387   the NHK medium supplier assured that specific lots of medium were available

388   to the laboratories. The test methods should be easily transferable to laboratories

389   experienced with mammalian cell culture methods.

390   •   Much of the training and expertise needed to perform the 3T3 and NHK NRU

391   test methods are common to all mammalian cell culturists. Additional technical

392   training would not be intensive since there are no extraordinary techniques

393   needed and these test methods are similar in general performance to other *in*

394   *vitro* mammalian cell culture assays. GLP training should be provided to

395   technicians to ensure proper adherence to protocols and documentation

396   procedures.

397   •   Price levels for commercial testing for one chemical are $1120 to $1850 (**Table**

398   **11-2**) for *in vitro* NRU cytotoxicity testing to determine the IC_{50} (IIVS, personal

399   communication) versus $750 - $3750 (**Table 11-2**) for *in vivo* rat acute oral

400   testing for LD_{50} determination. Comparison of costs of the *in vitro* testing to *in*

401   *vivo* testing is difficult since the *in vitro* NRU cytotoxicity test methods are not

402   replacements for the animal testing. Use of these test methods may not

403   necessarily reduce the overall cost of the *in vivo* rat acute oral toxicity test but

404   can reduce the number of animals needed for a study.

405

406

407

1 **12.0   REFERENCES**
2
3 Adegunloye BI, Sofola OA. 1997. Differential responses of rat aorta and mesenteric artery to
4 norepinephrine and serotonin *in vitro*. Pharmacology 55:25-31.
5
6 Ames A III. 2000. CNS energy metabolism as related to function. Brain Res Rev. 34:42-68.
7
8 Anon. 1993. Annexes I, II, III, and IV to Commission Directive 93/21/EEC of 27 April 1993
9 adapting to technical progress for the 18th time Council Directive 67/548/EEC on the
10 approximation of laws, regulations and administrative provisions relating to the
11 classification, packaging and labeling of dangerous substances. Official Journal of the
12 European Communities L110A, 1-86. (EU toxicity categories 25, 25-200, 200-2000, > 2000).
13
14 ASTM. 1999. American Society of Testing and Materials (ASTM) Standard E691-99,
15 Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a
16 Test Method. American National Standards Institute. Washington, D.C. Available:
17 http://www.ansi.org/ [accessed 22 December 2005].
18
19 ATSDR. 1999a. Toxicological Profile for Cadmium. Update PB/99/166621. US Dept. of
20 Health and Human Services, Atlanta GA. Available: http://www.atsdr.cdc.gov/toxpro2.html.
21 [accessed 2 June 2005].
22
23 ATSDR. 1999b. Toxicological Profile for Diethyl phthalate. PB/95/264214/AS. US Dept. of
24 Health and Human Services, Atlanta GA. Available: http://www.atsdr.cdc.gov/toxpro2.html.
25 [accessed 2 June 2005].
26
27 ATSDR. 2000a. Toxicological Profile for Endosulfan. .Update. PB/2000/108023. US Dept.
28 of Health and Human Services, Atlanta GA. Available:
29 http://www.atsdr.cdc.gov/toxpro2.html. [accessed 2 June 2005].
30
31 ATSDR. 2000b. Toxicological Profile for Chromium. Update. PB/2000/108022. US Dept. of
32 Health and Human Services, Atlanta GA. Available: http://www.atsdr.cdc.gov/toxpro2.html.
33 [accessed 2 June 2005].
34
35 ATSDR. 2001. Toxicological Profile for Di-n-butyl phthalate. Update. PB/2001/109104/AS.
36 US Dept. of Health and Human Services, Atlanta GA. Available:
37 http://www.atsdr.cdc.gov/toxpro2.html. [accessed 2 June 2005].
38
39 ATSDR. 2004a. Toxicological Profile for Pyrethrins and Pyrethroids. Update. PB2004-
40 100004. US Dept. of Health and Human Services, Atlanta GA. Available:
41 http://www.atsdr.cdc.gov/toxpro2.html. [accessed 2 June 2005].
42
43 ATSDR. 2004b. Toxicological Profile for Selenium. Update. PB2004-100005. US Dept. of
44 Health and Human Services, Atlanta GA. Available: http://www.atsdr.cdc.gov/toxpro2.html.
45 [accessed 2 June 2005].
46

47   Babich H, Martin-Alguacil N, Raul C, Rosenberg DW, Borenfreund E. 1991. Response of
48   human cell cultures to cytotoxicants requiring metabolic activation. In: Alternative Methods
49   in Toxicology (Goldberg AM, ed.), Vol. 8. New York:Mary Ann Liebert, Inc., 263-276.
50
51   Balls M, Botham PA, Bruner LH, Spielmann H. 1995. The EC/HO international validation
52   study on alternatives to the Draize eye irritation test for classification and labelling of
53   chemicals. Toxicol In Vitro 9**:**871-929.
54
55   Bernson V, Bondesson I, Ekwall B, Stenberg K, Walum E. 1987. A Multicentre evaluation
56   study of in vitro cytotoxicity. Altern Lab Anim. 14:144-146.
57
58   Bliss CI. 1938. The determination of the dosage-mortality curve from small numbers.
59   Quarterly Journal of Pharmacy and Pharmacology 11:192-214.
60
61   Bondesson I, Ekwall B, Hellberg S, Romert L, Stenberg K, Walum E. 1989. MEIC - A New
62   International Multicenter Project to Evaluate the Relevance to Human Toxicity of In Vitro
63   Cytotoxicity Tests. Cell Biol Toxicol. NOV 5(3):331-347.
64
65   Borenfreund E, Puerner J. 1984. A simple quantitative procedure using monolayer cultures for
66   cytotoxicity assays (HTD/ NR-90). J Tissue Culture Meth 9:7-9.
67
68   Borenfreund E, Puerner JA. 1985. Toxicity determination *in vitro* by morphological
69   alterations and neutral red absorption. Toxicol Lett. 24:119-124.
70
71   Botham PA. 2004. The validation of in vitro methods for skin irritation. Toxicol Lett.
72   149:387-390.
73
74   Brantom PG, Bruner LH, Chamberlain M, De Silva O, Dupuis J, Earl LK, et al. 1997. A
75   summary report of the COLIPA international validation study on alternatives to the Draize
76   rabbit eye irritation test. Toxicol In Vitro 11:141-179.
77
78   Canadian Centre for Occupational Health and Safety. 2005. IPCS INTOX Database.
79   Available: http://www.intox.org/databank/index.htm. [accessed 29 November 2005].
80
81   Casarett LJ, Klaassen CD, Doull J, eds. 2001. Casarett and Doull's Toxicology, The Basic
82   Science of Poisons. 6th Edition. New York:McGraw-Hill.
83
84   Charles River Laboratories Catalog. 2002. Available: http://www.criver.com. [accessed 2
85   June 2005].
86
87   Clemedson C, McFarlane-Abdulla E, Andersson M, Barile FA, Calleja MC, Chesné C, et al.
88   1996a. MEIC Evaluation of Acute Systemic Toxicity. Part I. Methodology of 68 *in vitro*
89   toxicity assays used to test the first 30 reference chemicals. Altern Lab Anim. 24 (suppl.
90   1):51-272.
91

92   Clemedson C, McFarlane-Abdulla E, Andersson M, Barile FA, Calleja MC, Chesné C, et al.
93   1996b. MEIC Evaluation of Acute Systemic Toxicity. Part II. *In vitro* results from 68 toxicity
94   assays used to test the first 30 reference chemicals and a comparative cytotoxicity analysis.
95   Altern Lab Anim. 24 (suppl. 1):273-311.
96
97   Clemedson C, Nordin-Andersson M, Bjerregaard H.F, Clausen J, Forsby A, Gustafsson H,
98   Hansson U, Isomaa B, Jorgensen C, et al. 2002. Development of an in vitro test battery for
99   the estimation of acute human systemic toxicity: an outline of the EDIT project Altern Lab
100 Anim. 30:313-321.
101
102 Clemedson, C. 2005. A-Cute-Tox project an Integrated Project under the EU 6FP with the
103 aim to optimize and pre-validate an *in vitro* test strategy for predicting human acute toxicity.
104 European Society of Toxicology In Vitro Newsletter. November. No. 18. p6.
105
106 Clothier R, Willshaw A, Cox H, Garle M, Bowler H, Combes R. 1999. The use of human
107 keratinocytes in the EU/COLIPA international in vitro phototoxicity test validation study and
108 the ECVAM/COLIPA study on UV filter chemicals. Altern Lab Anim. 27 (2):247-259 Mar-
109 Apr.
110
111 Coeck S, Balls M, Bowe G, Davis J, Gstraunthaler G, Hartung T, et al. 2005. Guidance on
112 Good Cell Culture Proactice. A Report of the Second ECVAM Task Force on Good Cell
113 Culture Practice. Altern Lab Anim. 33:261-287.
114
115 Coldwell BB, Boyd EM. 1966. The acute rectal toxicity of acetylsalicylic acid. Canadian
116 Journal of Physiology and Pharmacology 44: 909-918
117
118 Cooper-Hannan R, Harbell J, Coecke S, et al. 1999. The principles of good laboratory
119 practice: application to *in vitro* toxicology studies. Altern Lab Anim. 27:539-577.
120
121 Cosmetic Ingredient Review Panel. 1983. Final report on the safety assessment of sodium
122 borate and boric acid. J Am Coll Toxicol 2:87-125.
123
124 CPSC. 2003. Commercial Practices. Hazardous Substances and Articles; Administration And
125 Enforcement Regulations. 16 CFR 1500.42.
126
127 Creppy EE, Chiarappa P, Baudrimont I, Borracci P, Moukha S, Carratu MR. 2004.
128 Synergistic effects of fumonisin B1 and ochratoxin A: are in vitro cytotoxicity data
129 predictive of in vivo acute toxicity? Toxicology Sep 1;201(1-3):115-23.
130
131 Curren R, Bruner L, Goldberg A, Walum E. 1998. 13th meeting of the Scientific Group on
132 Methodologies for the Safety Evaluation of Chemicals (SGOMSEC): Validation and acute
133 toxicity testing. Environ Health Perspect. 106: (Suppl. 2). 419-425.
134
135 Curren RD, Moyer GO, Wilt N, Clear ML, Sizemore AM, Mun G. 2003. Assessment of
136 protocol variables in cytotoxicity assays utilizing BALB/c 3T3 cells and normal human

137   keratinocytes [Abstract]. Toxicologist 72:157. Available:
138   http://www.iivs.org/pages/publication_view.php?doc_id=43. [accessed 23 September 2005].
139
140   DOT. 2003. Shippers--General Requirements for Shipments and Packagings. Class 6,
141   Division 6.1 − Definitions. 49 CFR 173.132. Available:
142   http://www.gpoaccess.gov/cfr/index.html. [accessed 2 June 2005].
143
144   DOT. 2003. Shippers--General Requirements for Shipments and Packagings. Assignment of
145   Packing Group and Hazard Zones for Division 6.1 Materials. 49 CFR 173.133. Available:
146   http://www.gpoaccess.gov/cfr/index.html. [accessed 2 June 2005].
147
148   DOT. 1999. Code of Federal Regulations Title 49. SubChapter C. Assignment of packing
149   group and hazard zones for Division 6.1 materials.
150
151   Deichmann WB. 1969. Toxicology of Drugs and Chemicals. New York. Academic Press,
152   Inc. pg.67.
153
154   Dixon WJ, Massey FJ. 1981. Introduction to Statistical Analysis, 4th ed. Milwaukee:Quality
155   Press.
156
157   Ekwall B. 1983. Screening of Toxic Compounds in Mammalian Cell Cultures. Ann. New
158   York Acad. Sci. 407:64-77.
159
160   Ekwall B, Clemedson C, Craaford B, Ekwall B, Hallander S, Walum E, Bondesson I. 1998a.
161   MEIC evaluation of acute systemic toxicity. Part V. Rodent and human toxicity data for the
162   50 reference chemicals. Altern Lab Anim. 26 (suppl. 2):571-616.
163
164   Ekwall B, Clemedson C, Craaford B, Ekwall B, Hallander S, Walum E, Bondesson I. 1998b.
165   MEIC evaluation of acute systemic toxicity. Part VI. The prediction of human toxicity by
166   rodent LD50 values and results from 61 *in vitro* methods. Altern Lab Anim. 26 (suppl.
167   2):617-658.
168
169   Ekwall B, Clemedson C, Ekwall B, Ring P, Romert L. 1999. EDIT: A new international
170   multicentre programme to develop and evaluate batteries on *in vitro* tests for acute and
171   chronic systemic toxicity. Altern Lab Anim. 27:339-349.
172
173   Ekwall B, Ekwall B, Sjostrom M. 2000. MEIC evaluation of acute systemic toxicity: Part
174   VIII. Multivariate partial least squares evaluation, including the selection of a battery cell
175   line tests with a good prediction of human acute lethal peak blood concentrations for 50
176   chemicals. Altern Lab Anim. 28 (suppl. 1):201-234.
177
178   EPA. 1992. Reregistration Eligibility Document Citric Acid. List D, Case 4204. Office of
179   Pesticide Programs. Washington, DC:U.S. Environmental Protection Agency. Available:
180   http://cfpub.epa.gov/pesticides/reregistration. [accessed 2 June 2005].
181

182   EPA. 1995. Reregistration Eligibility Decision (RED) Diquat Dibromide. EPA 738-R-95-
183   016. Office of Prevention, Pesticides and Toxic Substances. Washington, DC:U.S.
184   Environmental Protection Agency. Available:
185   http://cfpub.epa.gov/pesticides/reregistration.[accessed 2 June 2005].
186
187   EPA. 1996. Trichloroacetic acid (CASRN 76-03-9) file in Integrated Risk Information
188   System. Office of Research and Development. Washington, DC:U.S. Environmental
189   Protection Agency. Available: http://www.epa.gov/iris.[accessed 23 September 2005].
190
191   EPA 738-R-95-016. Office of Prevention, Pesticides and Toxic Substances. Washington,
192   DC:U.S. Environmental Protection Agency. Available:
193   http://cfpub.epa.gov/pesticides/reregistration.[accessed 2 June 2005].
194
195   EPA. 1998. Product Properties Test Guidelines OPPTS 830.7840 Water Solubility: Column
196   Elution Method; Shake Flask Method. EPA 712–C–98–041. Washington, DC: U.S.
197   Environmental Protection Agency.
198
199   EPA. 2000. High Production Challenge Program. Office of Pollution Prevention and Toxics
200   (OPPT). Washington, DC:U.S. Environmental Protection Agency. Available:
201   http://www.epa.gov/chemrtk/volchall.htm [accessed 21 December 2005].
202
203   EPA. 2000a. Toxic Substances Control Guidelines; Final Rule. Code of Federal Regulations,
204   Title 40, Part 799. Available: http://www.epa.gov/oppt/chemtest/notice7.pdf. [accessed 2
205   June 2005].
206
207   EPA. 2000b. Toxicological Review of Chloral Hydrate. EPA/635/R-00/006. Washington,
208   DC:U.S. Environmental Protection Agency. Available: www.epa.gov/iris. [accessed 2 June
209   2005].
210
211   EPA. 2002a. Health Effects Test Guidelines OPPTS 870.1100 Acute Oral Toxicity. EPA
212   712–C–02–190. Washington, DC: U.S. Environmental Protection Agency.
213
214   EPA. 2002b. Statements of Policies and Interpretations. Toxicological and Ecological
215   Studies. 40 CFR 159.165. Available: http://www.gpoaccess.gov/cfr/index.html. [accessed 2
216   June 2005].
217
218   EPA. 2002c. Health Effects Test Guidelines OPPTS 870.1000 Acute Toxicity Testing -
219   Background. EPA 712–C–02–189. Washington, DC: U.S. Environmental Protection Agency.
220
221   EPA. 2002d. Emergency Planning and Notification. 40 CFR 355. Available:
222   http://www.access.gpo.gov/nara/cfr/waisidx_00/40cfr355_00.html. [accessed 2 June 2005].
223
224   EPA. 2003a. Good Laboratory Practice Standards. Toxic Substances Control Act. 40 CFR
225   792. Available: http://www.access.gpo.gov/nara/cfr/waisidx_03/40cfr160_03.html [accessed
226   21 December 2005].
227

228   EPA. 2003b. Good Laboratory Practice Standards. Federal Insecticide, Fungicide, and
229   Rodenticide Act. 40 CFR 160.
230
231   EPA. 2004. U.S. EPA HPV Challenge Program Submission. Propanoic acid, 2-hydroxy-,
232   compd. with 3-[2-(dimethylamino)ethyl] 1-(2-ethylhexyl) (4-methyl-1,3-
233   phenylene)bis[carbamate] (1:1) [CAS No. 68227-46-3]. Submitted by PPG Industries, Inc.
234   Revised December 17, 2004. http://www.**epa**.gov/chemrtk/prop2hyd/c13863rt3.pdf
235   [accessed 06 December 2005].
236
237   European Centre for the Validation of Alternative Methods. (ECVAM). 1997. Statement on
238   the scientific validity of the 3T3 NRU PT test (an in vitro test for phototoxic potential).
239   http://ecvam.jrc.it/index.htm. [accessed 15 August 2005].
240
241   EU. 2003. Report from the Commission to the Council and the European Parliament. Third
242   Report from the commission to the Council and the European Parliament of the statistics on
243   the number of animals used for experimental and other scientific purposes in the member
244   states of the European Union. Brussels, 22.01.2003, COM (2003) 19 final.
245
246   Fautrel A, Chesné C, Guillouzo A, De Sousa G, Placidi M, Rahmani R, et al. 1993. A
247   multicentre study of acute *in vitro* cytotoxicity in rat hepatocytes: tentative correlation
248   between *in vitro* toxicity and *in vivo* data. Altern Lab Anim. 21:281-284.
249
250   FDA. 1993. Toxicological Principles for the Safety Assessment of Direct Food Additives and
251   Color Additives Used in Food. Center for Food Safety and Applied Nutrition. Washington,
252   DC: U.S. Food and Drug Administration.
253
254   FDA. 2003. Good Laboratory Practice for Nonclinical Laboratory Studies. 21 CFR Part 58.
255   Available: http://www.access.gpo.gov/nara/cfr/waisidx_03/21cfr58_03.html [accessed 21
256   December 2005].
257
258   Fentem J, Fry J, Garle M, Gülden M, Seibert H, Voss J-U, et al. 1993. An International
259   Evaluation of Selected *In Vitro* Toxicity Test Systems for Predicting Acute Systemic
260   Toxicity. A report prepared for DGXI, CEC; Contract Numbers B92/B4-3063/14086 &
261   B92/B4-3040/14087. FRAME, Nottingham.
262
263   Fentem JH, Archer GEB, Balls M, Botham PA, Curren RD, Earl LK, et al. 1998. The
264   ECVAM international validation study on in vitro tests for skin corrosivity. 2. Results and
265   evaluation by the Management Team. Toxicol In Vitro. 12:483-524.
266
267   Fentem JH, Briggs D, Chesne C, Elliott GR, Harbell JW, Heylings JR, et al. 2001. A
268   prevalidation study on in vitro tests for acute skin irritation: results and evaluation by the
269   Management Team. Toxicol In Vitro. FEB 15(1):57-93.
270
271   Freshney RI. 2000. Culture of Animal Cells: A Manual of Basic Technique, 4th edition. New
272   York:Wiley-Liss.
273

274   Garle M., Fentem JH, Fry JR. 1994. *In Vitro* Cytotoxicity Tests for the Prediction of Acute
275   Toxicity *In Vivo*. Toxicol In Vitro. 8:1303-1312.
276
277   Gennari A, van den Berghe C, Casati S, Castell J, Clemedson C, Coecke S, et al. 2004.
278   Strategies t Replace In Vivo Acute Systemic Toxicity Testing. The Report and
279   Recommendations of ECVAM Workshop 50. Altern Lab Anim. 32:437-459.
280
281   Gettings SD, Bagley DM, Demetrulias JL, Dipasquale LC, Hintze KL, Rozen MG, et al.
282   1991. The CTFA evaluation alternatives program: An evaluation of *in vitro* alternatives to
283   the Draize primary eye irritation test. (Phase I) Hydro-alcoholic Formulations; (Part 2) Data
284   analysis and biological significance. In Vitro Toxicol. 4:247-288.
285
286   Gettings SD, Dipasquale LC, Bagley DM, Casterton PL, Chudkowski M, Curren RD, et al.
287   1994a. The CTFA evaluation of alternatives program: An evaluation of *in vitro* alternatives
288   to the Draize primary eye irritation test. (Phase II) Oil/water emulsions. Food Chem Toxicol.
289   32:943-976.
290
291   Gettings SD, Hintze KL, Bagley DM, Casterton PL, Chudkowski M, Curren RD, et al.
292   1994b. The CTFA evaluation of alternatives program: Phase III (surfactant-based
293   formulations). World Congress on Alternatives and Animal Use in the Life Sciences,
294   Baltimore, Maryland, USA, Nov. 14-19, 1993. In Vitro Toxicol. 7(2):166.
295
296   Gettings SD, Lordo RA, Hintze KL, Bagley DM, Casterton PL, Chudkowski M, et al. 1996.
297   The CFTA evaluation of alternatives program: An evaluation of *in vitro* alternatives to the
298   Draize Primary Eye Irritation Test. (Phase III) Surfactant-based Formulations. Food Chem
299   Toxicol. 34:79-117.
300
301   Gilman AG, Goodman LS, Rall TW, Murad F (eds.). 1985. Goodman and Gilman's The
302   Pharmacological Basis of Therapeutics, 7th Edition. New York:Macmillan Publishing Co.
303
304   Glaxo Wellcome. 2000. Myleran,(busulfan) Product Information. Research Triangle Park,
305   NC:Glaxo Wellcome, Inc.
306
307   GraphPad Software. 1994-2000. Prism® version 3.0a for Macintosh. GraphPad Software,
308   San Diego, California. Available: www.graphpad.com. [accessed 21 November 2005].
309
310   Grisham JW, Smith GJ. 1984. Predictive and Mechanistic Evaluation of Toxic Responses in
311   Mammalian Cell Culture Systems. Pharmacol Rev. 36(Suppl.):151S-171S.
312
313   Guzzie PJ. 1994. Lethality Testing. In: *In Vitro Toxicology*. (S.C. Gad, ed). Raven Press,
314   New York. pp.57-86.
315
316   Hackenberg U, Bartling H. 1959. Messen und Rechnen im pharmakologischen Laboratorium
317   mit einem speziellen Zahlensystem (WL24-System). Arch Exp Pathol Pharmakol 235:437-
318   463.
319

320 Halle W, Liebsch M, Traue D, Spielmann H. 1997. Reduktion der Tierzahlen bei der
321 Einstufung von Stoffen in die EU-Toxizitätsklassen fur akute orale Toxizität mit Hilfe von
322 Daten aus dem Register der Zytotoxizität (RC). Alternativen zu Tierexperimenten ALTEX
323 14:8-15.
324
325 Halle W. 1998. Toxizitätsprüfungen in Zellkulturen für eine Vorhersage der akuten Toxizität
326 (LD_{50}) zur Einsparung von Tierversuchen. Life Sciences/ Lebenswissenschaften, Volume 1,
327 94 pp., Jülich: Forschungszentrum Jülich. English translation: Halle W. 2003. The Registry
328 of Cytotoxicity: Toxicity testing in cell cultures to predict acute toxicity (LD50) and to
329 reduce testing in animals. Altern Lab Anim. 31:89-198.
330
331 Hall W. and Spielmann H. 1992. Two procedures for the prediction of acute toxicity (LD50)
332 from cytotoxicity data. Altern Lab Anim. 20:40-49.
333
334 Hardman JG, Limbird LE, Molinoff PB, Ruddon RW, Gilman AG, eds. 1996. Goodman and
335 Gilman's The Pharmacological Basis of Therapeutics, 9th Edition. New York:McGraw-Hill.
336
337 Harlan Company. 2002. Available: http://www.harlan.com/us/index.htm [accessed 2 June
338 2005]
339
340 Harbell JW, Koontz SW, Lewis RW, Lovell D, Acosta D. 1997. IRAG working group 4. Cell
341 cytotoxicity assays. Interagency Regulatory Alternatives Group. Food Chem Toxicol. 35:79-
342 126.
343
344 Hartung T, Balls M, Bardouille C, Blanck O, Coecke S, Gstrauthaler G, and Lewis D. 2002.
345 Good Cell Culture Practice: ECVAM Good Cell Culture Practice Task Force Report 1.
346 Altern Lab Anim. 30:407-414. Available: http://ecvam.jrc.it/publication/index5007.html
347 [accessed 02 September 2005].
348
349 Heimann R, Rice RH. 1983. Rat Esophageal and Epidermal-Keratinocytes - Intrinsic
350 Differences in Culture and Derivation of Continuous Lines. J Cell Physiol 117 (3):362-367.
351
352 Hintze KL, Janus J, Marenus KD, Muscatiello MJ, Pape WJW, Renskers KJ, et al. 1992.
353 Development of potential alternatives to the Draize eye test - The CTFA evaluation of
354 alternatives program - Phase II - Review of materials and methods. Altern Lab Anim.
355 20:164-171.
356
357 Hunter WJ, Lingk W, Recht JP. 1979. Intercomparison study on the determination of single
358 administration toxicity in rats. Journal of the Association of Official Analytical Chemists
359 62:864-873.
360
361 ICCVAM. 2000. The Revised Up-and-Down Procedure: A Test Method for Determining the
362 Acute Oral Toxicity of Chemicals and Products. Proposed Test Method and Background
363 Review Document, April 14, 2000. National Institute for Environmental Health Sciences,
364 Research Triangle Park, NC. Available: http://iccvam.niehs.nih.gov/. [accessed 2 June
365 2005].

366
367   ICCVAM. 2001a. Report Of The International Workshop On In Vitro Methods For
368   Assessing Acute Systemic Toxicity. NIH Publication No. 01-4499. National Institute for
369   Environmental Health Sciences, Research Triangle Park, NC. Available:
370   http://iccvam.niehs.nih.gov/. [accessed 2 June 2005].
371
372   ICCVAM. 2001b. Guidance Document On Using In Vitro Data To Estimate In Vivo Starting
373   Doses For Acute Toxicity. NIH Publication No. 01-4500. National Institute for
374   Environmental Health Sciences, Research Triangle Park, NC. Available:
375   http://iccvam.niehs.nih.gov/. [accessed 2 June 2005].
376
377   ICCVAM. 2001c. The Revised Up-and-Down Procedure: A Test Method for Determining
378   the Acute Oral Toxicity of Chemicals. NIH Publication No. 02-4501. National Institute for
379   Environmental Health Sciences, Research Triangle Park, NC. Available:
380   http://iccvam.niehs.nih.gov/. [accessed 2 June 2005].
381
382   ICCVAM. 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised,
383   and Alternative Test Methods. NIH Publication No. 03-4508. National Institute for
384   Environmental Health Sciences, Research Triangle Park, NC. Available:
385   http://iccvam.niehs.nih.gov/. [accessed 2 June 2005].
386
387   Institute of Medicine. 2004. Forging a Poison Prevention and Control System. Washington:
388   National Academies Press.
389
390   INVITTOX. 1991. The FRAME Modified Neutral Red Uptake Cytotoxicity Test. Protocol
391   Number 3 (IP-3A). http://embryo.ib.amwaw.edu.pl/invittox/. [accessed 12 Sep 2005].
392
393   Karczmar A. 1998. Invited review: Anticholinesterases: dramatic aspects of their use and
394   misuse. Neurochem Int. **32**:401-411.
395
396   King AV, Jones PA. 2003. In-house assessment of a modified in vitro cytotoxicity assay for
397   higher throughput estimation of acute toxicity. Toxicol In Vitro. Oct-Dec;17(5-6):717-22.
398
399   Knox P, Uphill PF, Fry JR, Benford J, Balls M. 1986. The FRAME Multicentre Project on *In*
400   *Vitro* Cytotoxicity. Food Chem Toxicol. 24:457-463.
401
402   Lipnick RL, Cotruvo JA, Hill RN, Bruce RD, Stitzel KA, Walker AP, et al. 1995.
403   Comparison Of The Up-And-Down, Conventional Ld(50), And Fixed-Dose Acute Toxicity
404   Procedures. Food Chem Toxicol. Mar 33:(3):223-231.
405
406   Litchfield JT, Wilcoxon F. 1949. A Simplified Method of Evaluating Dose-Effect
407   Experiments. J Pharmacol Exp Therap 96:99-113.
408
409   Litovitz TL, Klein-Schwartz W, White S, et al. 2000. 1999 Annual Report of the American
410   Association of Poison Control Centers Toxic Exposure Surveillance System. Am J Emerg
411   Med 18 (5):517-574 SEP.

412

413   The MathWorks. 1996-2004. MATLAB® software. The MathWorks, Inc., Natick,
414   Massachusetts.

415

416   Matsuno T, Masuda, Shimizu. 1971. Pipobroman: acute toxicity and general toxicological
417   reaction. Kiso to Rinsho Clinical Report 5:1894-1917.

418

419   MDL Information Systems 2001, 2002. Registry of Toxic Effects of Chemical Substances
420   (RTECS®) accessed online by subscription.

421

422   Microsoft Corporation. 1998-2001. Microsoft® Excel® 2000 software for Mac®. Microsoft
423   Corporation, Redmond, Washington.

424

425   Miller LC, Tainter ML. 1944. Estimation of the LD50 and its error by means of logarithmic-
426   probit graph paper. Proc Soc Exp Biol Med 57:261-264.

427

428   National Toxicology Program. 2002. Annual Plan For Fiscal Year 2002. Public Health
429   Service, Department of Health and Human Services. NIH Publication No. 03-5309. National
430   Institute for Environmental Health Sciences, Research Triangle Park, NC. Available:
431   http://ntp.niehs.nih.gov/index.cfm?objectid=03611496-026C-BB0F-B2A1527014B2D4B1.
432   [accessed 2 June 2005].

433

434   National Library of Medicine (NLM). 2005. Haz-Map, Occupational Exposure to Hazardous
435   Agents, Specialized Information Services, National Institutes of Health, Department of
436   Health & Human Services. Available: http://hazmap.nlm.nih.gov/. [accessed 29 November
437   2005].

438

439   National Library of Medicine. 2000, 2001, 2002, 2005. Hazardous Substances Data Bank.
440   Accessed online via TOXNET. National Institutes of Health, Department of Health &
441   Human Services. Available: http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB. [accessed
442   2 June 2005].

443

444   Occupational Safety and Health Act. 2003. Toxic and Hazardous Substances. Hazard
445   Communication 29 CFR 1910.1200
446   Available: http://www.gpoaccess.gov/cfr/index.html. [accessed 2 June 2005].

447

448   Occupational Safety and Health Act. 2000. Labeling Requirements for Pesticides and
449   Devices. Labeling Requirements. 40 CFR 156.10. Available:
450   http://www.gpoaccess.gov/cfr/index.html. [accessed 2 June 2005].

451

452   OECD. 1996. Guideline for Testing of Chemicals, 423, Acute Toxic Class Method. Paris
453   France: OECD.

454

455   OECD. 1998. OECD Series on Principles of Good Laboratory Practice and Compliance
456   Monitoring Number 1: OECD principles on Good Laboratory Practice. (as revised in 1997).

457   ENV/MC/CHEM(98)17. Paris: OECD. Available: http://www.oecd.org [accessed 2 June
458   2004].
459
460   OECD. 1998a. Harmonized Integrated Hazard Classification System for Human Health and
461   Environmental Effects of Chemical Substances as Endorsed by the 28th Joint Meeting of the
462   Chemicals Committee and the Working Party on Chemicals in November 1998, Part 2, p. 11.
463
464   OECD. 1998b. OECD Guideline for Testing of Chemicals 425: Acute oral toxicity—Up-and-
465   Down Procedure. OECD, Paris.
466
467   OECD. 2000. Guidance Document on the Recognition, Assessment and Use of Clinical
468   Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation.
469   Environmental Health and Safety Monograph Series on Testing and Assessment No. 19.
470   Paris France: OECD. Available:
471   [http://www.oecd.org/document/22/0,2340,en_2649_34377_1916054_1_1_1_1,00.html]
472
473   OECD. 2001a. Guideline for Testing of Chemicals, 425, Acute Oral Toxicity – Up-and-
474   Down Procedure. Paris France: OECD. Available: http://www.oecd.org [accessed 2 June
475   2004].
476
477   OECD. 2001b. OECD Series on Testing and Assessment, 33, Harmonized Integrated
478   Classification System for Human Health and Environmental Hazards of Chemical Substances
479   and Mixtures. ENV/JM/MONO(2001)6. Paris France: OECD. Available:
480   http://www.oecd.org [accessed 12 September 2004].
481
482   OECD. 2001c. Guideline for Testing of Chemicals, 420, Acute Oral Toxicity – Fixed Dose
483   Method. Paris France: OECD.
484
485   OECD. 2001d. Guideline For Testing of Chemicals, 423, Acute Oral Toxicity – Acute Toxic
486   Class Method. Paris France: OECD.
487
488   OECD. 2004a. OECD Series on Principles of Good Laboratory Practice and Compliance
489   Monitoring Number 14: Advisory Document of the Working Group on Good Laboratory
490   Practice: The Application of the Principles of GLP to In Vitro Studies. ENV/JM/MONO
491   (2004)26. Paris: OECD.
492
493   OECD. 2004b. Guideline for Testing of Chemicals, 432, *In Vitro* 3T3 NRU Phototoxicity
494   Test. Paris France: OECD. Available: http://www.oecd.org. [accessed 30 August 2005].
495
496   Orphan Medical, Inc. 1999. NDA 20-954 Busulfex™ (busulfan) Injection. NDA 20-954.
497   Available: www.fda.gov/cder/foi/label/1999/20954lbl.pdf. [accessed 2 June 2005].
498
499   Pesticide Action Network North America. 2005. Pesticide Action Network [PAN] Pesticides
500   Database. Available: http://www.pesticideinfo.org/Index.html. [accessed 29 November 2005]
501

502   Peloux AF, Federici C, Bichet N, Gouy D, Cano JP. 1992. Hepatocytes in primary culture: an
503   alternative to LD_{50} testing? Altern Lab Anim. 20:8-26.
504
505   Phillips JC, Gibson WB, Yam J, Alden CL, Hard GC. 1990. Survey of the QSAR and *In*
506   *Vitro* Approaches for Developing Non-animal Methods to Supersede the *In Vivo* LD50 Test.
507   Food Chem Toxicol. 28(5):375-394.
508
509   Rasmussen ES. 1999 Cytotoxicity of MEIC Chemicals Nos. 11-30 in 3T3 Mouse Fibroblasts
510   with and without Microsomal Activation. In Vitr Mol Toxicol. 12(3):125-132.
511
512   Riddell RJ, Panacer DS, Wilde SM, Clothier RH, Balls M. 1986. The importance of exposure
513   period and cell type in *in vitro* cytotoxicity tests. Altern Lab Anim. 14:86-92.
514
515   Roguet R. Colovio J, Gaetani Q., Sossou KG, Rogier A. 1993. Cytotoxicity of 28 MEIC
516   chemicals to rat hepatocytes using two viability endpoints: correlation with acute toxicity
517   data in rat and man. Altern Lab Anim. 21:216-224.
518
519   SAS Institute Inc. 1999. SAS/STAT User's Guide Version 8. Cary, NC:SAS Institute, Inc.
520
521   Schlede E, Mischke U, Roll R, Kayser D. 1992. A national validation study of the acute-
522   toxic-class method--an alternative to the LD50 test. Archives of Toxicology **66**:455-470.
523
524   Schlede E, Mischke U, Diener W, Kayser D. 1995. The international study of the acute toxic
525   class method (oral). Archives of Toxicology **69**:659-670.
526
527   Schmahl D, Osswald H. 1970. Experimental studies on the carcinogenic effects of anticancer
528   chemotherapeutics and immunosuppressive agents. Arzneimittelforschung. Oct;20(10):1461-
529   1467.
530
531   Seibert H, Gulden M, Kolossa M, Schepers G. 1992. Evaluation of the Relevance of Selected
532   In Vitro Toxicity Test Systems for Acute Systemic Toxicity. Altern Lab Anim. APR
533   20(2):240-245
534
535   Seibert H, Balls M, Fentem JH, Bianchi V, Clothier RH, Dierickx PJ, et al. 1996. Acute
536   Toxicity Testing *In Vitro* and the Classification and Labelling of Chemicals. The Report and
537   Recommendations of ECVAM Workshop 16. Altern Lab Anim. 24:499-510.
538
539   Sina JF, Galer DM, Sussman RG, Gautheron PD, Sargent EV, Leong B, et al. 1995. A
540   collaborative evaluation of seven alternatives to the Draize eye irritation test using
541   pharmaceutical intermediates. Fundam Appl Toxicol. 26:20-31.
542
543   Spielmann H, Genschow E, Liebsch M, Halle W. 1999. Determination of the starting dose
544   for acute oral toxicity (LD_{50}) testing in the up and down procedure (UDP) from cytotoxicity
545   data. Altern Lab Anim. **27**:957-966.
546

547   Spielmann H, Balls M, Dupuis J, Pape WJW, Pechovitch G, de Silva O, et al. 1998.
548   EU/COLIPA "*In Vitro P*hototoxicity" Validation Study, Results of Phase II (blind trial), Part
549   1: the 3T3 NRU phototoxicity test. Toxicol In Vitro. 12:305-327.
550
551   Spielmann H, Balls M, Dupuis J, Pape WJW, De Silva O, Holzhütter H.-G, et al. 1998b. A
552   study on the phototoxic potential of UV filter chemicals from Annex VII of EU Directive
553   76/768/EEC in the 3T3 NRU *in vitro* phototoxicity test. Altern Lab Anim. 26**:**679-705.
554
555   Spielmann H, Liebsch M, Kalweit S, Moldenhauer F, Wirnsberger T, Holzhütter HG, et al.
556   1996. Results of a validation study in Germany on two *in vitro* alternatives to the Draize eye
557   irritation test, the HET-CAM test and the 3T3 NRU cytotoxicity test. Altern Lab Anim.
558   24**:**741-858.
559
560   Spielmann H, Kalweit S, Liebsch M, Wirnsberger T, Gerner I, Bertram-Neis E, et al. 1993.
561   Validation study of alternatives to the Draize eye irritation test in Germany: Cytotoxicity
562   testing and HET-CAM test with 136 industrial chemicals. Toxicol In Vitro. 7**(4)**:505-510.
563
564   Spielmann H, Gerner S, Kalweit S, Moog R, Wirnserberger T, Krauser K, et al. 1991.
565   Interlaboratory assessment of alternatives to the Draize eye irritation test in Germany.
566   Toxicol In Vitro. 5:539-542.
567
568   Stallard N, Whitehead A. 2004. A statistical evaluation of the fixed dose procedure. Altern
569   Lab Anim. Sep;32 Suppl 2:13-21.
570
571   Taconic Farms, Inc. 2002. Taconic Farms Animal Models, Sprague Dawley® Outbred Rats.
572   Available: http://www.taconic.com/anmodels/spragued.htm. [accessed 2 June 2005].
573
574   Thompson WR. 1947. Use of moving averages and interpolation to estimate median-
575   effective dose. Bacteriol Rev 11:115-145.
576
577   Thomson PDR®. 2004. Physicians' Desk Reference, 58th edition. Montvale, NJ: Thomson
578   PDR.
579
580   Triglia D, Wegener PT, Harbell J, Wallace K, Mathesen D, Shopsis C. 1989. Interlaboratory
581   validation study of the keratinocyte neutral red bioassay from Clonetics Corporation. In: In
582   Vitro Toxicology, New Directions (Goldberg A, ed). Alternative Methods in Toxicology,
583   Volume 7. New York: Mary Ann Liebert, 357-365.
584
585   UN. 1977. United Nations Economic and Social Council. Joint meeting of the RID safety
586   committee and the group of experts on the transportation of dangerous goods. Trans/GE 15/R
587   274, 2.
588
589   UN. 2005. Globally Harmonized System of Classification and Labelling of Chemicals
590   (GHS), First Revised Edition. [ST/SG/AC.10/30/Rev.1]. United Nations, New York and
591   Geneva. Available: http://www.unece.org/trans/danger/publi/ghs/ghs_rev01/01files_e.html
592   [accessed 12 September 2005].

593   Vernot EH, MacEwen JD, Haun CC, Kinkead ER. 1977. Acute toxicity and skin corrosion
594   data for some organic and inorganic compounds and aqueous solutions. Toxicol Appl
595   Pharmacol. 42:417-423.
596
597   Wallum, E. 1998. Acute Oral Toxicity. Environ Health Perspect. 106:497-503.
598
599   Weil CS. 1952. Tables for convenient calculation of median effective dose (LD50 or ED50)
600   and instructions in their use. Biometrics 8(3):249-263.
601
602   Weil CS, Wright GJ. 1967. Intra- and Interlaboratory Comparative Evaluation of Single Oral
603   Test. Toxicol Appl Pharmacol. 11:378-388.
604
605   WHO. 1998. The WHO Recommended Classification of Pesticides by Hazard and
606   Guidelines to Classification 1998-1999. International Programme of Chemical Safety.
607   WHO/PCS/98.21. Geneva:World Health Organization.
608
609   Willshaw A, Moore LJ, Balls M. 1994. *In vitro* alternatives for the detection of photoirritant
610   chemicals - the EEC COLIPA trial. Toxicol In Vitro. 8:723-725.
611
612   Zbinden G, Flury-Roversi M. 1981. Significance of the LD50-test for the toxicological
613   evaluation of chemical substances. Arch Toxicol 47(2):77-99.
614
615   Zuang V, Balls M, Botham PA, Coquett A, Corsini E, Curren RD, et al. 2002. Follow-up to
616   the ECVAM prevalidation study on in vitro tests for acute skin irritation. Altern Lab Anim.
617   30:109-129.
618
619
620
621

1   **13.0   GLOSSARY[1]**

2

3   **Accuracy[2]:** (a) The closeness of agreement between a test method result and an accepted

4   reference value. (b) The proportion of correct outcomes of a test method. It is a measure of

5   test method performance and one aspect of "relevance". Accuracy is highly dependent on

6   the prevalence of positives in the population being examined.

7

8   **Acute Toxic Class (ATC) method:** An acute oral systemic toxicity test method based on

9   testing groups of animals at fixed doses in a sequential manner. The lethality outcomes are

10 used to classify a test substance into the appropriate GHS acute oral toxicity category.

11

12 **Adjusted R^2:** R^2 values that are adjusted for the relative proportion of data points to

13 explanatory variables. Adjusted $R^2 = 1 - (1 - R^2)[(n - 1)/(n - k - 1)]$ where k = number of

14 independent variables and n = number of observations. See "coefficient of determination."

15

16 **ANOVA:** One-way (and two-way) analysis of variance. ANOVA compares the

17 measurements (continuous variables) of three or more groups when the data are categorized

18 in one way (one-way) or two ways (two-way). ANOVA assumes that the populations

19 compared are normally distributed and that the variances for the groups to be compared are

20 approximately equal.

21

22 **Assay[2]:** The experimental system used. Often used interchangeably with "test" and "test

23 method."

24

25 **Biphasic dose-response:** Dose-response in which cytotoxicity increases (as dose increases),

26 plateaus, and then increases again. See **Section 2.6.3**.

27

---

[1] The definitions in this Glossary are restricted to their uses with respect to *in vitro* cytotoxicity testing and the NRU test methods.
[2] Definition used by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM 2003).

28   **Category prediction:** The GHS hazard category that includes the predicted LD_{50} value for a

29   test chemical.

30

31   **Coded substances:** Substances labeled by code rather than name so that they can be tested

32   and evaluated without knowledge of their identity or anticipation of test results. Coded

33   substances are used to avoid intentional or unintentional bias when evaluating laboratory or

34   test method performance.

35

36   **Coefficient of variation:** A statistical representation of the precision of a test. It is expressed

37   as a percentage and is calculated as follows:

38

$$\left(\frac{standard\ deviation}{mean}\right)\ \times\ 100\%$$

39

40

41   **Coefficient of determination:** In linear regression, it denotes the proportion of the variance

42   in Y and X that is shared. Its value ranges between zero and one and it is commonly called

43   called "R^2." For example, $R^2 = 0.45$, indicates that 45% of the variance in Y can be

44   explained by the variation in X and that 45% of the variance in X can be explained by the

45   variation in Y.

46

47   **Concordance[2]:** The proportion of all substances tested that are correctly classified as

48   positive or negative. It is a measure of test method performance and one aspect of

49   "relevance." The term is often used interchangeably with "accuracy." Concordance is

50   highly dependent on the prevalence of positives in the population being examined. In the

51   NICEATM/ECVAM study, concordance was used to describe the proportion of test

52   substances that were correctly classified into GHS acute oral toxicity hazard categories, or to

53   describe the proportion of test substances for which the laboratories obtained the same

54   classification result.

55

56   **Confluency:** A state in which cells in culture come into contact with other cells in the same

57   culture to form a complete sheet of cells (monolayer). For this study, confluency is

58   determined as a percentage of cell coverage of the tissue culture vessel growth surface (e.g.,

59   cell monolayer has 80% confluency).

60

61   **Cytotoxicity:** The adverse effects resulting from interference with structures and/or

62   processes essential for cell survival, proliferation, and/or function. For most chemicals,

63   toxicity is a consequence of non-specific alterations in "basal cell functions" (i.e., via

64   mitochondria, plasma membrane integrity, etc.), which may then lead to effects on organ-

65   specific functions and/or death of the organism. These effects may involve the integrity of

66   membranes and the cytoskeleton, cellular metabolism, the synthesis and degradation or

67   release of cellular constituents or products, ion regulation, and cell division.

68

69   **Definitive test:** The main test of the cytotoxicity assay for determining the IC_{50}. The

70   concentration closest to the range finder test IC_{50} serves as the midpoint of the concentrations

71   tested in a definitive test. Compared to the range finder test, the definitive test uses a smaller

72   dilution factor for the concentrations tested.

73

74   **Discordant chemicals:** Chemicals for which the LD_{50} is not accurately predicted by the IC_{50}

75   (and the associated regression formula) or the GHS toxicity category is not accurately

76   predicted by the IC_{50} (and the associated regression formula). Also referred to as "outliers."

77

78   **EDIT:** Evaluation-guided Development of New *In vitro* Test Batteries. An international

79   project coordinated by the Scandinavian Society for Cell Culture to develop new *in vitro* tests

80   for toxicity and toxicokinetics to be incorporated into test batteries for predicting acute and

81   chronic systemic toxicity.

82

83   **Endpoint[2]:** The biological process, response, or effect assessed by a test method.

84

85   **Fixed Dose Procedure (FDP):** An acute oral systemic toxicity test method based on testing

86   groups of animals at fixed doses. Evident toxicity outcomes are used to classify a test

87   substance into the appropriate GHS acute oral toxicity category.

88

89   **F_G:** An empirical factor for the RC regression line that represents the expected precision of

90   LD_{50} predictions from basal cytotoxicity data. The LD_{50} values of 73% of the 347 RC

91   chemicals are localized in the dose range around the RC regression line by $F_G \leq \log 5$. The

92   factor represents the expected difference between the LD_{50} determined in animal experiments

93   and the LD_{50} estimated from the IC_{50} on the RC regression line.

94

95   **Geometric mean:** The antilog of the mean of the logarithm of the values. It is less affected

96   by extreme values than the arithmetic mean.

97

98   **Globally Harmonized System (GHS):** A classification system presented by the United

99   Nations that provides (a) a harmonized criteria for classifying substances and mixtures

100 according to their health, environmental and physical hazards, and (b) a harmonized hazard

101 communication elements, including requirements for labeling and safety data sheets.

102

103 **Good Laboratory Practices (GLP)[2]:** Regulations promulgated by the U.S. Food and Drug

104 Administration and the U.S. Environmental Protection Agency, and principles and

105 procedures adopted by the Organization for Economic Cooperation and Development and

106 Japanese authorities that describe record keeping and quality assurance procedures for

107 laboratory records that will be the basis for data submissions to national regulatory agencies.

108

109 **Guidance Document:** *Guidance Document on Using In Vitro Data to Estimate In Vivo*

110 *Starting Doses for Acute Toxicity* (ICCVAM 2001b).

111

112 **Hazard[2]:** The potential for an adverse health or ecological effect. A hazard potential results

113 only if an exposure occurs that leads to the possibility of an adverse effect being manifested.

114

115   **Hill function:** The IC_{50} values are determined from the concentration-response using a Hill

116   function which is a four parameter logistic mathematical model relating the concentration of

117   the test chemical to the response (typically following a sigmoidal shape).

118    $$Y = \text{Bottom} + \frac{\text{Top} - \text{Bottom}}{1 + 10^{(\text{logIC50}-X)\text{HillSlope}}}$$

119   where Y= response, X is the logarithm of dose (or concentration), Bottom is the minimum

120   response, Top is the maximum response, $logIC_{50}$ is logarithm of X at the response midway

121   between Top and Bottom, and HillSlope describes the steepness of the curve.

122

123   **Hill function (revised):** Some unusual dose-responses did not fit the Hill function well. To

124   obtain a better model fit, the Bottom parameter was estimated without constraints (the

125   previous practice was to use Bottom = 0). However, when Bottom $\neq$ 0, the EC_{50} reported by

126   the Hill function was not the same as the IC_{50} since the Hill function relies on EC_{50} defined

127   as the point midway between Top and Bottom. Thus, the Hill function calculation using the

128   Prism® software was rearranged to calculate the concentration corresponding to the IC_{50} as

129   follows.

130

$$X = \log EC_{50} - \frac{\log\left(\frac{Top - Bottom}{Y - Bottom} - 1\right)}{HillSlope}$$

131

132   X is the logarithm of concentration at 50% response, $logEC_{50}$ is logarithm of concentration at

133   the response midway between Top and Bottom, Top is the maximum response, Bottom is the

134   minimum response, Y = 50 (i.e., 50% response), and HillSlope describes the steepness of the

135   curve.

136

137   **IC_{50}:** test chemical concentration producing 50% inhibition of the endpoint measured (i.e.,

138   cell viability).

139

140   **Interlaboratory reproducibility[2]:** A measure of whether different qualified laboratories

141   using the same protocol and test substances can produce qualitatively and quantitatively

142   similar results. Interlaboratory reproducibility is determined during the prevalidation and

143   validation processes and indicates the extent to which a test method can be transferred

144   successfully among laboratories.

145

146   **Intralaboratory repeatability[2]:** The closeness of agreement between test results obtained

147   within a single laboratory when the procedure is performed on the same substance under

148   identical conditions within a given time period.

149

150   **Intralaboratory reproducibility[2]:** The first stage of validation; a determination of whether

151   qualified people within the same laboratory can successfully replicate results using a specific

152   test protocol at different times.

153

154   *In vitro:* In glass. Refers to assays that are carried out in an artificial system (e.g., in a test

155   tube or petri dish) and typically use single-cell organisms, cultured cells, cell-free extracts, or

156   purified cellular components.

157

158   *In vivo:* In the living organism. Refers to assays performed in multicellular organisms.

159

160   **K_{ow}**: Octanol:water partition coefficient.

161

162   **LC_{50}:** Acute lethal serum or blood concentrations.

163

164   **LD_{50}**: The calculated value of the oral dose that produces lethality in 50% of test animals

165   (rats and mice). The LD_{50} values serve as reference values for the *in vitro* tests.

166

167   **LD_{50} (initial):** Acute oral rat and mouse LD_{50} values used during the chemical selection

168   process. For RC chemicals, LD_{50} values were those used in the RC regression, which were

169   largely from the 1983/84 RTECS®. For chemicals that were not included in the RC, the

170   initial LD_{50} values came from HSDB or 2002 RTECS®.

171

172   **LD$_{50}$ (reference):** Acute oral rodent LD$_{50}$ values from rats and mice were located through

173   literature searches and references from major toxicity databases such as RTECS®. Studies

174   were reviewed to identify the most appropriate LD$_{50}$ values for each chemical. Values

175   obtained using feral animals, preanesthetized animals, or animals less than 4 weeks of age

176   were not used. Values reported as inequalities were not used. Reference LD$_{50}$ values were

177   determined by calculating the geometric mean of the acceptable LD$_{50}$ values. Data were

178   used in generation of the 3T3 and NHK NRU regressions.

179

180   **Maximum:minimum value**: Ratio of minimum acceptable LD$_{50}$ to maximum acceptable

181   LD$_{50}$.

182

183   **MEIC:** Multicentre Evaluation of *In Vitro* Cytotoxicity. An international effort established

184   by the Scandinavian Society for Cell Toxicology and initiated in 1983 to evaluate the

185   relationship and relevance of *in vitro* cytotoxicity for predicting the acute toxicity of

186   chemicals in humans.

187

188   **Millimolar regressions:** Linear regressions with IC$_{50}$ values in mmol/L and LD$_{50}$ values in

189   mmol/kg.

190

191   **Negative control:** An untreated sample containing all components of a test system, except

192   the test substance solvent, which is replaced with a known non-reactive material, such as

193   water. This sample is processed with test substance-treated samples and other control

194   samples to determine whether the solvent interacts with the test system.

195

196   **Neutral red (NR):** A weakly cationic water-soluble dye that stains living cells by readily

197   diffusing through the plasma membrane and concentrating in lysosomes where it

198   electrostatically binds to the anionic lysosomal matrix.

199

200   **Neutral red uptake (NRU):** Concentration of neutral red dye in the lysosomes of living

201   cells. Altering the cell surface or the lysosomal membrane by a toxicological agent causes

202   lysosomal fragility and other adverse changes that gradually become irreversible. The NRU

203   test method makes it possible to distinguish between viable, damaged, or dead cells because

204   these changes result in decreased uptake and binding of NR measurable by optical density

205   absorption readings in a spectrophotometer.

206

207   **NHK: N**ormal **H**uman epidermal **K**eratinocytes (from neonatal foreskin).

208

209   **Optical density (OD):** The absorption (i.e., OD measurement) of the resulting colored

210   solution (colorimetric endpoint) in the NRU assay measured at 540 nm ± 10 nm in a

211   spectrophotometric microtiter plate reader using blanks as a reference

212

213   **Outlier:** For any measurement, an extreme value in the NICEATM/ECVAM study was

214   referred to as an "outlier" if it passes a statistical test for outliers at the 99% level. With

215   respect to chemicals, it refers to chemicals that do not fit (using the specified criteria) an

216   IC_{50}-LD_{50} linear regression model. It may also refer to chemicals for which the predicted

217   GHS toxicity category does not match the reference *in vivo* GHS toxicity category.

218

219   **Performance[2]:** The accuracy and reliability characteristics of a test method (see "accuracy",

220   "reliability").

221

222   **pH:** A measure of the acidity or alkalinity of a solution. pH 7.0 is neutral; higher pHs are

223   alkaline, lower pHs are acidic.

224

225   **Plate reader:** A spectrophotometric device for measuring light intensity as a function of

226   color/wavelength (i.e., optical density/absorption at 540 nm ± 10 nm for NRU) in 96-well

227   microtiter tissue culture plates.

228

229   **Positive control:** A sample containing all components of a test system and treated with a

230   substance known to induce a positive response, which is processed with the test substance-

231   treated and other control samples to demonstrate the sensitivity of each experiment and to

232   allow for an assessment of variability in the conduct of the assay over time.

233

234 **Predictivity**[2]: Proportion of *in vivo* category matches for all substances with *in vitro*

235 predictions for a particular category. Predictivity is an indicator of test accuracy.

236

237 **Protocol**[2]**:** The precise, step-by-step description of a test, including the listing of all

238 necessary reagents, criteria and procedures for the evaluation of the test data.

239

240 **Quality assurance (QA)**[2]**:** A management process by which adherence to laboratory testing

241 standards, requirements, and record keeping procedures is assessed independently by

242 individuals other than those performing the testing.

243

244 **Range finder:** Initial test performed to determine starting doses for the main (definitive) test.

245 The NRU assays test eight concentrations of the test chemical or the positive control (PC) by

246 diluting the stock solution in log dilutions to cover a large concentration range.

247

248 **RC regression:** $\log (LD_{50}) = 0.435 \times \log (IC_{50}) + 0.625$; for estimating an LD_{50} value in

249 mmol/kg (body weight) from an IC_{50} value (in mM).

250

251 **Reduction alternative**[2]**:** A new or modified test method that reduces the number of animals

252 required.

253

254 **Reference substances:** Substances selected for use during the research, development,

255 prevalidation, and validation of a proposed test method because their response in the *in vivo*

256 reference test method or the species of interest is known (see "reference test"). Reference

257 substances should represent the classes of chemicals for which the proposed test method is

258 expected to be used and cover the range of expected responses (negative, weak to strong

259 positive).

260

261 **Reference test method**[2]**:** The accepted *in vivo* test method used for regulatory purposes to

262 evaluate the potential of a test substance to be hazardous to the species of interest.

263

264   **Refinement alternative[2]:** A new or modified test method that refines procedures to lessen or

265   eliminate pain or distress in animals or enhances animal well-being.

266

267   **Registry of Cytotoxicity (RC):** Database that consists of *in vivo* acute oral toxicity data (i.e.,

268   LD_{50} values) from rats and mice and *in vitro* cytotoxicity data (i.e., IC_{50} values) from

269   multiple cell lines and cytotoxicity endpoints for 347 chemicals with known molecular

270   weights (Halle 1998). A regression model constructed from these data was proposed by

271   ZEBET, as a method to reduce animal use by identifying the most appropriate starting doses

272   for acute oral systemic toxicity tests

273

274   **Relevance[2]:** The extent to which a test method correctly predicts or measures the biological

275   effect of interest in humans or another species of interest. Relevance incorporates

276   consideration of the "accuracy" or "concordance" of a test method.

277

278   **Reliability[2]:** A measure of the degree to which a test method can be performed reproducibly

279   within and among laboratories over time. It is assessed by calculating intra- and inter-

280   laboratory reproducibility and intralaboratory repeatability.

281

282   **Replacement alternative[2]:** A new or modified test method that replaces animals with

283   nonanimal systems or one animal species with a phylogenetically lower one (e.g., a mammal

284   with an invertebrate).

285

286   **Reproducibility[2]:** The consistency of individual test results obtained in a single laboratory

287   (intralaboratory reproducibility) or in different laboratories (interlaboratory reproducibility)

288   using the same protocol and test substances (see intra- and inter-laboratory reproducibility).

289

290   **RTECS®:** Registry of Toxic Effects for Chemical Substances**.** Compendium of data

291   extracted from the open scientific literature. The database includes toxicity data (e.g., acute

292   toxicity) and specific numeric toxicity values (e.g., LD_{50}). Compiled by the U.S. National

293   Institute for Occupational Safety and Health (NIOSH) and now licensed to MDL Information

294   Systems, Inc.

295   **Sensitivity[2]:** The proportion of all positive substances that are classified correctly as positive

296   in a test method. It is a measure of test method accuracy.

297

298   **Simulation modeling:** Computer simulation modeling of the acute systemic toxicity assays

299   to determine animal use. The simulation process uses a simulated population of animals for

300   testing, a reference endpoint (i.e., "true" LD_{50} value), and its assumed log-normal

301   distribution. Morality is assumed to have a mean equal to the log of the true LD_{50}. The SD,

302   which reflects the variability of the simulated population, is the inverse of the slope of the

303   dose-mortality curve. Due to a lack of information for the real dose-mortality curve, the

304   simulations assumed slopes of 0.5, 0.8, 2, 4, and 8.3.

305

306   **Solubility:** The amount of a test substance that can be dissolved (or thoroughly mixed with)

307   culture medium or solvent. The solubility protocol was based on a U.S. EPA guideline (EPA

308   1998) that involves testing for solubility in a particular solvent, beginning at a relatively high

309   concentration and proceeding to successively lower concentrations by adding more solvent as

310   necessary for dissolution. Testing stops when, upon visual observation, the procedure

311   produces a clear solution with no cloudiness or precipitate.

312

313   **Solvent control:** An untreated sample containing all components of a test system, including

314   the solvent that is processed with the test substance-treated and other control samples to

315   establish the baseline response for the samples treated with the test substance dissolved in the

316   same solvent. When tested with a concurrent negative control, this sample also demonstrates

317   whether the solvent interacts with the test system.

318

319   **Specificity[2]:** The proportion of all negative substances that are classified correctly as

320   negative in a test method. It is a measure of test method accuracy.

321

322   **Spirit of GLP:** Guidance provided in the Statement of Work specifically for the non GLP-

323   compliant laboratory that participated in the validation study. Based on the GLP standards

324   referenced in the ECVAM Workshop 37 Report (Cooper-Hannan 1999) and the OECD

325   Principles of GLP (OECD 1998). *"Laboratories that are non GLP-compliant shall adhere to*

326   *GLP principles and other method parameters. Documentation and accountability shall be*

327   *equal to GLP requirements. Laboratories must make assurances that they are equal in*

328   *performance criteria and that there is parity amongst the laboratories."*

329

330   **TESS**: Toxic Exposure Surveillance System. A comprehensive poisoning surveillance

331   database maintained by the American Association of Poison Control Centers (AAPCC).

332

333   **Test[2]:** The experimental system used; used interchangeably with "test method" and "assay".

334

335   **Test method[2]:** A process or procedure used to obtain information on the characteristics of a

336   substance or agent. Toxicological test methods generate information regarding the ability of

337   a substance or agent to produce a specified biological effect under specified conditions.

338   Used interchangeably with "test" and "assay". See also "validated test method" and

339   "reference test".

340

341   **Test method component:** Structural, functional, and procedural elements of a test method

342   that are used to develop the test method protocol. These components include unique

343   characteristics of the test method, critical procedural details, and quality control measures.

344

345   **3T3:** BALB/c 3T3 clone A31 mouse fibroblasts developed in 1968 from disaggregated 14- to

346   17-day-old BALB/c mouse embryos (American Type Culture Collection [ATCC]; # CCL-

347   163).

348

349   **Tiered testing:** A testing strategy where all existing information on a test substance is

350   reviewed, in a specified order, before *in vivo* testing.

351

352   **Toxicity underpredicted:** Actual LD_{50} value of a test substance is lower than the predicted

353   LD_{50} value.

354

355   **Toxicity overpredicted:** Actual LD_{50} value of a test substance is higher than the predicted

356   LD_{50} value.

357   **Transferability[2]:** The ability of a test method or procedure to be accurately and reliably

358   performed in different, competent laboratories.

359

360   **Up-and-Down Procedure (UDP):** An acute oral systemic toxicity test method used to

361   minimize the number of animals required to estimate the acute oral toxicity of a chemical,

362   estimate the LD_{50} and confidence intervals (CI), and observe signs of toxicity. Single

363   animals are tested sequentially. Subsequent doses are based on the outcome of the previous

364   animal.

365

366   **Validated test method[2]:** An accepted test method for which validation studies have been

367   completed to determine the accuracy and reliability of this method for a specific proposed

368   use.

369

370   **Validation[2]:** The process by which the reliability and accuracy of a procedure are established

371   for a specific purpose.

372

373   **Vehicle control (VC):** The VC consists of appropriate cell culture medium for the cells in

374   the test (i.e., DMEM for 3T3 cells and keratinocyte growth medium for the NHK cells). For

375   chemicals dissolved in DMSO, the VC consists of medium with the same amount of solvent

376   as that used in the test chemical concentrations that are applied to the 96-well test plate. The

377   final DMSO concentration is ≤ 0.5 % (v/v) in the VCs.

378

379   **Volatility:** Ability of a test chemical to evaporate. A general indicator of volatility issues in

380   the NRU test methods is the percent difference in the mean OD values for the two VC

381   columns on the test plate. If the difference is greater than 15%, then chemical volatility can

382   be suspected, especially if the VC adjacent to the highest test concentration had a

383   significantly reduced OD value. Volatility may be an issue for compounds with a specific

384   gravity of less than 1.

385

386   **Weight of evidence (process):** The strengths and weaknesses of a collection of information

387   are used as the basis for a conclusion that may not be evident from the individual data.

388   **Weight regressions:** Linear regressions with IC_{50} values in µg/mL and LD_{50} values in

389   mg/kg.

390

391   **ZEBET:** The German National Center for the Documentation and Evaluation of Alternative

392   Methods to Animal Experiments.

393